

CC (3H)-pyrimidine-5-phosphate);
CC -rib-7 gene: HTP-reductase (transforming 2,5-diamino-6-ribosylamino-4-
CC (3H)-pyrimidine-5-phosphate into 2,5-diamino-ribitylamino-
CC 2,4-(1H,3H)-pyrimidine-5-phosphate (DRAP));
CC -rib-2 gene: (DRAP)-deaminase (formation of 5-amino-6-ribityl
CC amino-2,4-(1H,3H)-pyrimidinol);
CC -rib-4 gene: 6,7-dimethyl-8-ribityllumazine (DMRL)-synthase;
CC -rib-5 gene: riboflavin-synthase;
CC -rib-3 gene: L-3,4-dihydroxy-2-butanone-4-phosphate (DBP)-synthase.
SQ Sequence 591 AA;

Query Match 7.3%; Score 85.5; DB 1; Length 591;

Best Local Similarity 21.8%; Pred. No. 0.52; Mismatches 59; Indels 83; Gaps 11;

Matches 47; Conservative 27; Mismatches 59; Indels 83; Gaps 11;

QY 21 HRVERFSQATRRTPPOPTRVILFNKPYDVLVQPTDE-----AGR--- 63

DB 147 HKVHRH-----EPPVT-----SKPIDIV--FEDEDILVIXPSSIPVHPTGRYRF 189

QY 63 STLKDFIPVQ---GVYAAGRLDRDSEGLLVLTNDGVLQARL--TOPGKRTGKIYVQVEG 117

DB 190 NTITKMLERQLGVSVHPCNRLDPTSGLMFLAKTPLGADRMGDMQKAREVTKYVARVKG 249

QY 118 EPDASLAKLRNGVTNDGPTLPAGIERVNEPWLPRNPPI-----RERKS 164

DB 250 E-----PPGIVEVDKP--VRSVNPKNALNAVCEMSDENAKHAKT 287

QY 165 I-----PTSLWKITLYEGNRQVRRTAHVGF 192

DB 288 VFQVSYDQGTSLVCKEPTGTGRHQIRVHLQYLGF 323

RESULT 5

W93393

ID W93393 standard; Protein: 660 AA.

DT 11-JUN-1999 (first entry)

DE Human HEV ORF 2 protein from strain Hyderabad.

KW Swine hepatitis E virus; HEV; cross-reaction; antibody: human; therapy:

OS Hepatitis E virus.

PN W0904029-A2.

PF 28-JAN-1999.

PR 17-JUL-1997; U14665.

PA (USSR) US DEPT HEALTH & HUMAN SERVICES.

PI Emerson SU, Meng X, Purcell RH;

DR WPI; 99-132270/11.

PT New isolated swine hepatitis E virus - used to develop products for

PT the diagnosis, prevention and treatment of hepatitis E virus

PT infection in mammals, particularly humans

PS Example 1; Fig 3A; 70pp; English.

CC This invention describes a swine hepatitis E virus (HEV) and its natural

CC mutants which are capable of cross-reacting with antibodies reactive

CC with a human HEV strain or natural mutants. The HEV and the proteins

CC can be used in vaccines for immunising against HEV infection. The swine

CC HEV can be used in humans to prevent possible infection by human HEV. The

CC swine HEV can also be used as a therapeutic treatment for infection by

CC other strains of HEV. The swine HEV can also be used for the production

CC of antibodies which can be used in therapy, detection and diagnosis. The

CC products can also be used for determining the susceptibility of cells or

CC organs to infection with swine HEV. The swine HEV is particularly useful

CC for the development of agents for the prevention, treatment and detection

CC of human HEV because it is not a human virus and thus can be handled both

CC experimentally and clinically without fear of severe infection and/or

CC contamination.

SQ Sequence 660 AA;

Query Match

Best Local Similarity 7.2%; Score 84; DB 1; Length 660;

Matches 59; Conservative 24; Mismatches 99; Indels 58; Gaps 12;

QY 7 IPENTMTSTFRKHVERFS--ROATRR-----TPEOPTRVILFNKP-YDVLPOFTDE 59

DB 311 LTPGNTNT-----RVSRYSSTARHRLRGADGTAELTTAATRMKDLFTSTNGVGE 363

QY 60 AGR-----STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTOPGKRTGK 109

DB 364 IGRGIALTLFNADTLGLGLTELISAG-----GQLFYSRPVV-----SANGPTVK 411

QY 110 IYVVOVEGEPDDASLAKLRNGVTNDGPTLPAGIERVNEPWLPRNPPIRERKSPTS- 169

DB 412 L-YTSVENAQDQKGA-IPNDIDLGESRVVIOYDQNHQEDRPTSPAPSPFVLRAND 469

QY 169 --WLKITYEGNRQVRRTAHVGFPTLRIYAGSYTLDLSANG-----EWRDVT 218

DB 470 VLWLSLTAAYEQSTYSGSTGPV-----YVSDSVTLNVNATGAQAVARSLDWTKVT 520

RESULT 6

W80707

ID W80707 standard; Protein: 163 AA.

AC W80707;

DT 24-DEC-1998 (first entry)

DE S. pneumoniae protein of unknown function.

KW Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;

KW virulence; antibody; infection; detection; treatment; hypothetical;

OS Streptococcus pneumoniae.

PN W09826072-A1.

PD 18-JUN-1998.

PF 09-DEC-1997; U22578.

PR 13-DEC-1996; US-036281.

PA (ELIL) LILLY & CO ELI.

PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SK,

PI Mills BJ, Norris FH, Peery RB, Rostek PK, Rostek PR,

PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,

PI Young Bellido ML;

DR WPI; 98-348529/30.

DR N-P8DB; V65267.

PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips

PT for evaluating gene expression, and identification of virulence

PT genes

PS Claim 3; Page 299; 33pp; English.

CC This sequence represents a Streptococcus pneumoniae protein of unknown

CC function. The invention provides DNA sequences (V65201 to V65304) from

CC the Streptococcus pneumoniae genome and corresponding protein sequences

CC (W80605 to W80728). The protein sequences are classified as hypothetical,

CC cell wall biosynthetic, external target, or minimal gene set proteins. A

CC recombinant host containing a vector comprising any of the above nucleic

CC acids can be used for the recombinant expression of the proteins. The

CC invention also provides a DNA chip having arrayed on it at least 15 base

CC pair fragment of any one or more of these DNA sequences. The DNA chip can

CC be used methods for evaluating gene expression in S. pneumoniae and for

CC identifying virulence genes in S. pneumoniae. Antibodies that selectively

CC bind to the above proteins or peptide fragments can be used to treat

CC S. pneumoniae infection. The antibodies can also be used to detect

SQ Sequence 163 AA;

Query Match 7.1%; Score 83; DB 1; Length 163;

Best Local Similarity 23.6%; Pred. No. 0.15;

Matches 33; Conservative 27; Mismatches 56; Indels 24; Gaps 6;

QY 65 LKDFIPVQGVYAAG---RLDRDSEGLLVLTNDGVLQARLTQ---PGKRTGKIYVQVEGE- 119

DB 24 IKDLSINGVLRPIVHRIDKDTSGLLMTAKNDADHLVLAQELKDKRSURKRYWAIHVGNL 83

QY 119 PDDASLAKLRNGVTNDGPTLPAG-IERYNEPEWLPRNPPIRERKSI-----PTSLWKI 172

DB 84 PNDRGVIE-----APIGRSEKDKKQAVTAKGPAVTRFHVLERFDGYSILVEL 131

QY 173 TLYEGNRQVRRTAHVGF 192

CC A PCR product amplified using primers based on partial peptide
CC fragments of purified A.niger SOX was used to screen an A.niger
CC N400 genomic DNA library in lambda-EMBL4. A 2.5kb EcoRI fragment
CC comprising the entire SOX gene was identified in positive plaques
CC and was sequenced. Additional upstream sequence information was
CC obtained from a 6.4kb NsiI fragment also present in positive
CC plaques. The SOX protein is encoded by 3 exons.
CC Sequence 392 AA;

Query Match 7.0%; Score 81; DB 1; Length 392;
Best Local Similarity 25.5%; Pred. No. 0.89;
Matches 42; Conservative 28; Mismatches 63; Indels 32; Gaps 9;
QY 2 IMQLIPENTMTKTSFRKHVERESSRQATRTPEQPTRVILENKPVDLPQFTDEAG 61
DB 258 IFRVHFTDGSVPNTF-----ITNPTAQRSTLPEELSLVMVDNKIDT---TDYTG 306
QY 62 RSLKDFIPVQGYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGIYVQVEGEPDD 121
DB 307 MRT-----SLSGYAVG--DCNSDG-----STNVPHAMFS--GKAGVYVHVMSREESN 352
QY 122 ASLAKLRNGVTLNDGPTLPAGIRSV--NEPEWLPNRPPIREKRS 164
DB 353 AATSK-RD---FDRRALEKQTERVWGNEMEDLKWVLENNHRRS 392

RESULT 10

ID R14619 standard; Protein; 660 AA.
AC R14619;
DT 10-JAN-1992 (first entry)
DE Protein encoded by ORF 2 of Burmese ET-NANB viral strain.
KW Enterically transmitted non-A, non-B hepatitis virus; hepatitis C;
KW HCV; E.coli strain BB4; ATCC deposit number 67717; Burma.
OS Enterically transmitted non-A, non-B hepatitis virus.
PN W09115603-A.
PD 17-OCT-1991.
PF 03-APR-1991; U02368.
PR 05-APR-1990; US-505888.
PA (GENE-) GENELABS INC.
PA (USSH) US DEPT HEALTH & HUMAN.
PI Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;
PI Fry RE;
DR WPI; 91-325242/44.
DR N-PSDB; Q14412.
PT New viral proteins from non A-non-B hepatitis agent - used to
PT treat and prevent enterically-transmitted non-A non-B hepatitis
PT virus
PS Disclosure; Page 19; 117pp; English.
CC A positive clone ET1.1 was identified in a library prepared from
CC bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.
CC Both strands of ET1.1 were sequenced. Identity of the sequence with
CC sequences in etiologic agents has been confirmed by locating a
CC similar sequence in a viral strain isolated in Burma. This protein
CC is encoded by the second longest ORF (ORF 2) of the Burma strain.
CC (See Q14410 for ET1.1).
CC Sequence 660 AA;

Query Match 7.0%; Score 81; DB 1; Length 660;
Best Local Similarity 24.2%; Pred. No. 1.9;
Matches 58; Conservative 25; Mismatches 99; Indels 58; Gaps 12;
QY 7 ITPEWTMTKTSFRKHVERESS--RQATRR---TPEQPTRVILENKP-YDVLPTQFTDE 59
DB 311 LTPGNTNT-----RVSRSSSTARHRLRGADGTAELTTAATRFMKDLYFTSTNGVE 363
QY 60 AGR-----STLKDFIPVQGYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
DB 364 IGRGIALTFNLADTLGLVLPTELISAG-----QLFYSRPV-----SANGEPYVK 411
QY 110 IYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIRVNEPEWLPNRPPIREKRSIPT- 169

DB 412 L-YTSVENAQODKGIA-IPHDIDGESRVVYQDYDQHQEDRTPSPAPSRPSVLRAND 469
QY 169 --WKITLXEGNRNQVRRTAHYGFPTLRILRYAMGSYTLDSLNG-----EWRDVT 218
DB 470 VLWLSLTAAYEDQSTYGSSTGPPV-----YVSDSVILVNVATGAQAVARSILDWTKVT 520
RESULT 11
ID Y00238
AC Y00238 standard; Protein; 2032 AA.
DT 20-APR-1999 (first entry)
DE Enterococcus faecalis protein EF123.
KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW detection; attenuation; antigenic.
OS Enterococcus faecalis.
PN W09850554-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08959.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;
PI WPI; 99-070095/06.
DR N-PSDB; X20228.
PT New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection
PT Claim 9; Page 224-225; 301pp; English.
CC The present sequence represents a protein isolated from
CC Enterococcus faecalis. The present invention describes genes, proteins
CC and antigenic polypeptides isolated from E. faecalis. The proteins can
CC be used in vaccines for preventing or attenuating an infection caused
CC by a member of the Enterococcus genus in an animal. They can also be
CC used for detecting Enterococcus antibodies in a sample. The nucleotide
CC sequences can be used for detecting Enterococcus nucleic acids.
CC Products from the present invention can also be used for screening
CC compounds to identify agonists and antagonists of E. faecalis protein
CC activity.
CC Sequence 2032 AA;

Query Match 6.9%; Score 80; DB 1; Length 2032;
Best Local Similarity 24.5%; Pred. No. 13;
Matches 48; Conservative 19; Mismatches 83; Indels 46; Gaps 9;
QY 25 RFSRQNRTRTPPEQPTRVILENKPVDLPQFTDEAGRS---TLK-DFIPVQGYAAGRL 80
DB 1836 RFVETQA-----PSTYLLNETPSASFIAKDNOGKPAVYVLPKAPFNYQGAALKVKI 1887
QY 81 DRSEGL-----LVLNDGVLAARLTQPGKRTGIYVQVEGEPDD 121
DB 1888 DQKNALAGAEFKVDAETQGTVARSLRSDNQLGVNHLQPGKYT----FVETKA-PDG 1942
QY 122 ASLAK-----LRNGVTLNDGPTLPAGIRVNEPEWLPNRP---NPPIREKRSIPTSLWTKITL 174
DB 1943 YQLSKQAVAFITIAATKDKPELVNAGTFVNEKOPVSKKTKPNQPTTKQAARETCWLGL-- 2001
QY 175 YEGNRQVRRTAHVG 190
DB 2001 -PKTNTQVNYFFVFIG 2015
RESULT 12
ID Y00240
AC Y00240 standard; Protein; 2032 AA.
DT 20-APR-1999 (first entry)
DE Enterococcus faecalis protein EF124.
KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW detection; attenuation; antigenic.

100

QY 69 I-PVQGVTAAGRLDRDSEGLLVLTNDGVLA--RLTQPKRTGKIYYVQVEGPDASLA 125
| | | | | : : : | | | : | : : | : : | : : :
Db 706 IEFVQTQTVV-----DMKDVML-NDIIEQAAGRSRASDRGVSYYPTTDDDE----- 755
QY 126 IELRNGVTLNDGPT 138
| | | | |
Db 755 -----DGPT 758

RESULT 2
US-08-705-245-2
; Sequence 2, Application US/08705245
; Patent No. 6020189
; GENERAL INFORMATION:
; APPLICANT: Nathans et al., Jeremy
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR HOMOLOGOUS
; TITLE OF INVENTION: FACTORS (FHF's) AND METHODS OF USE
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/705,245
; FILING DATE: 30-AUG-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07265/094001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-50999
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 245 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-705-245-2

Query Match 7.0%; Score 82; DB 3; Length 245;
Best Local Similarity 26.7%; Pred. No. 0.1;
Matches 39; Conservative 19; Mismatches 44; Indels 44; Gaps 6;
QY 15 KTSPRKH-----RVERFSSRQATRRPEQPTVILFNKPYD-----VLPQF 56
| | | | | : : : | | | : | : : | : : | : : :
Db 35 KTSCKDKNLNVSFKLVGSKKRRRRPEQLKGIV--TKLISRQGVHLQADGTIDGT 92
| | | | | : : : | | | : | : : | : : | : : :
QY 57 TDEAGRSTLKDFIPV-----QGVYAAGRLDRDSEGLL-----VLTNDG 94
| | | | | : : : | | | : | : : | : : | : : :
Db 93 KDEDSTVTLNLPVGLRVVAIQGVQTKLYLANNSEGLYISELFTPECKESVPEY 178
| | | | | : : : | | | : | : : | : : | : : :
QY 95 VLQARLTQPKRTGKIYYVQV--EGE 118
| | | | | : : : | | | : | : : | : : | : : :
Db 153 VIYSSMIYRQQSGRGWYLGINKKE 178

RESULT 3
US-08-423-441-2
; Sequence 2, Application US/08423441
; Patent No. 5529926

GENERAL INFORMATION:
; APPLICANT: MAAT, JAN
; APPLICANT: MUSTERS, WOUTER
; APPLICANT: STAM, HEIN
; APPLICANT: SCHAAAP, PETER J.
; APPLICANT: VAN DE VONDERVOORT, PETER J.
; APPLICANT: VISSER, JACOB
; APPLICANT: VERBAKEL, JOHANNES M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA ENCODING A
; TITLE OF INVENTION: RIPENING FORM OF A POLYPEPTIDE HAVING SULFHYDRYL OXIDASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/423,441
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/044,620
; FILING DATE: 09-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 202744/T7019(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 248453CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-423-441-2

Query Match 7.0%; Score 81; DB 1; Length 392;
Best Local Similarity 25.5%; Pred. No. 0.28;
Matches 42; Conservative 28; Mismatches 63; Indels 32; Gaps 9;
QY 2 IMRQLITPENTMTKTSFRKHYRVERFSSRQATRRPEQPTVILFNKPYDVLPOFTDEAG 61
| | | | | : : : | | | : | : : | : : | : : :
Db 258 IFRVHTDSSVVPNTF-----ITNPTAORSLPEELSLVMVDNKIDT---TDYTG 306
| | | | | : : : | | | : | : : | : : | : : :
QY 62 RSTLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLAQRLTQPKRTGKIYYVQVEGPD 121
| | | | | : : : | | | : | : : | : : | : : :
Db 307 MRT-----SLSGVYAVG--DCNSDG-----STNVPHAFMS--GKRAGVYVHVMSRESN 352
| | | | | : : : | | | : | : : | : : | : : :
QY 122 ASLAKLRNGVTLNDGFTLPAGIERV--NEPELWLPNPIPERKS 164
| | | | | : : : | | | : | : : | : : | : : :
Db 353 AAISK-RD-----FDRRALEKQTERVMGVNEMEDLWKRVLNHHRS 392

RESULT 4
US-08-438-439C-2
; Sequence 2, Application US/08438439C
; Patent No. 5876967
; GENERAL INFORMATION:
; APPLICANT: Nathans, Jeremy
; APPLICANT: Smallwood, Phillip M.
; APPLICANT: Macke, Jennifer P.

```

RESULT      5
US-08-951-822-24
Sequence 24, Application US/08951822A
Patent No. 5989866
GENERAL INFORMATION:
APPLICANT: Delscher, Theresa A.
APPLICANT: Conklin, Darrell C.
APPLICANT: Raymond, Fencella
APPLICANT: Bukowski, Thomas R.
APPLICANT: Holderman, Susan D.
APPLICANT: Hansen, Birgit
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: NOVEL FGF HOMOLOGS
FILE REFERENCE: 96-20
CURRENT APPLICATION NUMBER: US/08/951,822A
CURRENT FILING DATE: 1997-10-16
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 24
LENGTH: 245
TYPE: PRT

```

```

; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 245 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-705-245-9

Query Match 6.8%; Score 79; DB 3; Length 245;
Best Local Similarity 26.7%; Pred. No. 0.23;
Matches 39; Conservative 18; Mismatches 45; Indels 44; Gaps 6

Qy 15 KTSRKH-----RVERSSQARTRPDPQTRVILENKPYD-----VLQPF 56
||| | :|: | :|: | :|: | :|: | :|: |
Db 35 KTSCKNKLNVFSRVKLGSKRRRRRPEPOLKGV--TKLYSRQYHLQQAQGTIDGT 92
||| | :|: | :|: | :|: | :|: | :|: |
Qy 57 TDEAGRSLTKDFIPV-----QGYAAAGRLDRDSEGLL-----VLTDNG 94
||| | :|: | :|: | :|: | :|: | :|: |

```


Db 93 KDEDSTYLFNLPVGLRVVAIQGVQTKLYLNMSEGYLTSHEHTPECKFKESVFENY 152
QY 95 VLQARLTQPKRTGKIYYVQV--EGE 118
Db 153 VTYSSMIYRQQSGRGWYLGKLNKEGE 178

RESULT 7

PC-US96-06665-2
Sequence 2, Application PC/TUS9606665
GENERAL INFORMATION:
APPLICANT: The Johns Hopkins University School of Medicine
TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR HOMOLOGOUS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06665
FILING DATE: 10-MAY-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.,
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/046W01
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 245 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US96-06665-2

Query Match 6.8%; Score 79; DB 4; Length 245;
Best Local Similarity 26.7%; Pred. No. 0.23;
Matches 39; Conservative 18; Mismatches 45; Indels 44; Gaps 6;

QY 15 KTSFRKH-----RVERFSSRQATRTPEQPQTRVILFNKPYD-----VLPQF 56
Db 35 KTSCKNKLNVFSRVKLGSKRRRRRPEQLKGV--TKLYSRQGYHLQLQADGTIDGT 92
QY 57 TDEAGRSTLKDFIPV-----QGYAAGRLDRDSEGLL-----VLTNDG 94
Db 93 KDEDSTYLFNLPVGLRVVAIQGVQTKLYLNMSEGYLTSHEHTPECKFKESVFENY 152
QY 95 VLQARLTQPKRTGKIYYVQV--EGE 118
Db 153 VTYSSMIYRQQSGRGWYLGKLNKEGE 178

RESULT 8

US-08-438-439C-11
Sequence 11, Application US/08438439C
Patent No. 5676967
GENERAL INFORMATION:
APPLICANT: Nathans, Jeremy
APPLICANT: Smallwood, Phillip M.
APPLICANT: Macke, Jennifer P.
TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR HOMOLOGOUS

TITLE OF INVENTION: FACTOR-2 AND METHODS OF USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/438,439C
FILING DATE: May 12, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/046001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-438-439C-11

Query Match 6.8%; Score 79; DB 2; Length 246;
Best Local Similarity 26.7%; Pred. No. 0.24;
Matches 39; Conservative 18; Mismatches 45; Indels 44; Gaps 6;

QY 15 KTSFRKH-----RVERFSSRQATRTPEQPQTRVILFNKPYD-----VLPQF 56
Db 35 KTSCKNKLNVFSRVKLGSKRRRRRPEQLKGV--TKLYSRQGYHLQLQADGTIDGT 92
QY 57 TDEAGRSTLKDFIPV-----QGYAAGRLDRDSEGLL-----VLTNDG 94
Db 93 KDEDSTYLFNLPVGLRVVAIQGVQTKLYLNMSEGYLTSHEHTPECKFKESVFENY 152
QY 95 VLQARLTQPKRTGKIYYVQV--EGE 118
Db 153 VTYSSMIYRQQSGRGWYLGKLNKEGE 178

RESULT 9

US-08-259-148A-17
Sequence 17, Application US/08259148A
Patent No. 5741490
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R.
APPLICANT: Bradley, Daniel W.
APPLICANT: Twu, Jr-Shin
APPLICANT: Purdy, Michael A.
APPLICANT: Tam, Albert W.
APPLICANT: Krawczynski, Krzysztof Z.
APPLICANT: Yarbough, Patrice D.
TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/259,148A
FILING DATE: 13-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 822,335
FILING DATE: 17-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0093.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 436 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9
PS-08-259-148A-17

```

Query Match      6.8%; Score 79; DB 1; Length 436;
Best Local Similarity 24.2%; pred. No. 0.56;
Matches 58; Conservative 25; Mismatches 99; Indels 58; Gaps 12;

QY  7  ITPENTMTKTSFRKRVVERFSS--RQATRR-----TPEQPTRVILFNKP-YDVLPOFTDE 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  87  LTPGNTNT-----RVSRYSTASHRRURGADGCTAELTTTAATRFMKDLYFTSTNGYGE 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  60  AGR-----STLKDFIPQGVYAAAGRLDRDSEGLLVLTNDGVQLQARLTQPKRGTK 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  140  IGRGIALTLFNLAOTLLGGLTELISSAG-----GQLFYSRPV-----SANGEPVTK 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  110  IYVVOVEGPDDASLAKRVNGVTNDGPTLPAGIERVNEPEWLWPRNPPIRKRISPT-S 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  188  L-YTSVENAAQDKGIA-IPHDIDLGESRWVIQDYDNQHQDRPTSPAPSRPFSVLRLND 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  169  --WLKIITLYEGENRQVRMTAHVGPPTLRLTRYAMGSTLDOSLANG-----EWRDVT 218
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  246  VLWLSLTAAYDQSTYGSSTGPV-----YVSDSVTLNVNATGAQAVARSLDWTKYT 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-08-484-054-17
; Sequence 17, Application US/08484054
; Patent No. 5770689
; GENERAL INFORMATION:
; APPLICANT: Reves, Gregory R.

```

APPLICANT: Bradley, Daniel W.
 APPLICANT: Twu, Jr-Shin
 APPLICANT: Purdy, Michael A.
 APPLICANT: Tam, Albert W.
 APPLICANT: Krawczynski, Krystof Z.
 TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dehlinger & Associates
 STREET: 350 Cambridge Avenue, Suite 250
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/484,054
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 870,985
 FILING DATE: 20-APRIL-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 822,335
 FILING DATE: 17-JAN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 505,888
 FILING DATE: 05-APRIL-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 420,921
 FILING DATE: 13-OCTOBER-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 367,486
 FILING DATE: 16-JUNE-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 336,672
 FILING DATE: 11-APRIL-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 208,997
 FILING DATE: 17-JUNE-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Sholtz, Charles K.
 REGISTRATION NUMBER: 38,615
 REFERENCE/DOCKET NUMBER: 4600-0093.38
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 324-0880
 TELEFAX: (415) 324-0960
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 436 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEetical: NO
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9
 PS-08-484-054-17

Query Match	6.88;	Score 79;	DB 1;	Length 436;
Best Local Similarity	24.28;	Pred. No. 0.56;		

Qy	7	ITPENTMTKTSFRKHVERESS--RQATR-----TPEQPQTRVILFNKP-YDVLPQFTDE	59
	:		
	:		
	:		
Db	87	LTPGNTNT-----RVSYRSSTARHLRGADCTALTTAATPFMKDLFTSTNGVGE	139
	:		
	:		
	:		
Qy	60	AGR-----STLKDFIPVQGVYAAGRLDRDSEGLVLVTNDGVQLARLTPOGKRGTG	109
	:		
	:		
	:		

```

140  IGRIALTFLNADTLGLGLTELISAG-----GQFYSPVV-----SANGEPYVK 187
QY  IYVQVEGEDDASLAKLNGVTLNDGPILPACIERVNEPEWLWPNRPPIRERKS IPTS- 169
    : ||| : : : : : : : : : : : : : : : : : : : : : : : :
188  L-TYSVENAQQDKGIA-IPHDIDLSGRVVYQIDYDQHQDRPTSPSPADSPFVSLRAND 245
    : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  --WLKITLYEGRNQVRMRMTAHVGFPTLRLIRYAMGSYTLDSISLNG-----EWRDVT 218
    || : || : || : || : || : || : || : || : || : || : || : ||
246  VLWLSLTAERYDYDOSTVGSSTGPV-----YVSODVTLVNVTAQAQAVARSLDWTKYT 296
    : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11

```

US-07-876-941A-17
: Sequence 17, Application US/07876941A
: Patent No. 5885768
:
: GENERAL INFORMATION:
:
: APPLICANT: Reyes, Gregory R.
: APPLICANT: Bradley, Daniel W.
: APPLICANT: Tam, Albert W.
: APPLICANT: Mitchell, Carl
:
: TITLE OF INVENTION: Hepatitis E Virus Rep
:
: TITLE OF INVENTION: Antibodies
:
: NUMBER OF SEQUENCES: 76
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Dehlinger & Associates
: STREET: 350 Cambridge Avenue, Suite 250
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306
:

```

ZAP. 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/876.941A
FILING DATE: 01-MAY-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 822,335
FILING DATE: 17-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,888
FILING DATE: 05-APRIL-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 420,921
FILING DATE: 13-OCTOBER-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 367,486
FILING DATE: 16-JUNE-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 336,672
FILING DATE: 11-APRIL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 208,997
FILING DATE: 17-JUNE-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0093.33
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 436 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9

US-07-876-941A-17

Query Match	6.8%	Score 79;	DB 2;	Length 436;
Best Local Similarity	24.2%;	Pred. No. 0.56;		
Matches	58;	Conservative 25;	Mismatches 99;	Indels 58; Gaps 12;
QY	7	ITPENTWTKTSFRKRRHYERESS--RQATRR-----TPPQPQTRVLFLNKP-YDVLPQPTDE	59	
Db	87	LTGCTNTW-----RVSRISSTARHLRGAAGTAEELTTAATRFMKDLYFTSTNGVGE	139	
QY	60	AGR-----STLKDFIPVOGVYAAGRLDRDSEGLLVLTNDGVLAQLRTPQGRKTGK	109	
Db	140	IGRGIALTLFNLAIDLTLGGFLPTELISSAG-----GQLFYSRPVV-----SANGEPYVK	187	
QY	110	IYVQVGEPEDDASLAKLRNGVLTNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIFTS-	169	
Db	188	L-YTSVENAQODKGIA-IPHDIDLGRSVVIQYDINQHQEDRTTSPSPAPRPFVSLRAND	245	
QY	169	--WLKTYLZGRNRQVRMTAHVGFPTLRLIRYAMGSYTLDSLNG-----EWRDVT	218	
Db	246	VZLWLSYAAEYDQSTYGSSTGPV-----YVSDSVTLNVATGAQAVARSILDMTKVT	296	

RESULT 12

PCR-US95-13703-27
Sequence 27, Application PC/TUS9513703
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND
TITLE OF INVENTION: USES THEREFOR
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13703
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0293.41
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Hepatitis E virus (Burma strain)
INDIVIDUAL ISOLATE: r62kDa, 56.5 kDa
PCT-US95-13703-27

Query Match 6.8%; Score 79; DB 4; Length 525;
Best Local Similarity 24.2%; Pred. No. 0.74;
Matches 58; Conservative 25; Mismatches 99; Indels

QY		110	IYYVQVEGEPPDASLAKLRNGVTLDGPTLPAGIERVNPEWLWPNPPTI	ERKSKIPTS-	169
D8		301	L-ITSVENAQDGIA-I PHDIOLGESRVVIQDYDNHQEQDRTPSPASRPF	SVLRAND	358
QY		169	--WLKITLYEGRNQVRMTAHGFPTLRILRYAMGSYTLDSLSANG	-----ENRDYT	218
D8		359	VLMISITAAEVDOSTYSGSSGTGPV----	YVSDSVILNVATCAQAVARSLDWT	409

```

RESULT 14
PCT-US95-13703-15
; Sequence 15, Application PC/TUS9513703
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND
; TITLE OF INVENTION: USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13703
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0293.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hepatitis E virus (Burma str)
; INDIVIDUAL ISOLATE: r62kda, FIGURE 4
PCT-US95-13703-15

```

Query Match 6.8%; Score 79; DB 4; Length 549;
Best Local Similarity 24.2%; Pred. NO. 0.8;
Matches 58; Conservative 25; Mismatches 99; Indels 58; Gaps 12

Qy	7	ITPENTWTKTSFKRHRVERESS--RQATRR-----TPEQPTRVILFNKP-YDVLPQFTDE	59
Db	200	LTPGNTNT-----RYSRYSSSTARHRLRRGADGTAELTTTAATRFMKDLYFTSTNGVGE	252
Qy	60	AGR-----STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVQLARLTQFGKRTGK	109
Db	253	IGRGIAULTFLNLADTLGGLPTELSSAG-----GQLFYSRPV-----SANGEPTVK	300
Qy	110	IYYVQVGEPPDASLAKLRNGVTLNDGFTLPAGLIERVNEPELWPRNPPTIRERKSIPTS-	169
Db	301	L-YTSVENAQDQKCIA-IPHDIDLGSRVVIQDYDNQHEQDRTPSPAPSRPSVFLRAND	358
Qy	169	--WLKITYEGRNQVRRMTAHGVFPTPLRLIRYAMGSYTLDSLANG-----EWRDVT	218

Db 359 VLWLSLTAAYDQSTYGSSTGPGV-----YVSDSVTLNVNATGAQAVARSLDWTXVT 409

RESULT 15

US-08-240-049B-15
; Sequence 15, Application US/08240049B
; Patent No. 5686239
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Tam, Albert W.
; APPLICANT: Yarbough, Patrice O.
; TITLE OF INVENTION: Hepatitis E Virus Peptides and Methods
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240.049B
; FILING DATE: 09-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles K. Sholtz
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0093.35; G32P4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hepatitis E Virus (Burma strain)
; INDIVIDUAL ISOLATE: ORF-2
US-08-240-049B-15

Query Match 6.8%; Score 79; DB 1; Length 660;
Best Local Similarity 24.2%; Pred. No. 1.1;
Matches 58; Conservative 25; Mismatches 99; Indels 58; Gaps 12;
QY 7 ITPENTMTKTSFRKHVERFSS--RQATRR-----TPEQPTRVILFNKP-YDVLFPQFTDE 59
Db 311 LTPGNTNT-----RVSRYSSTARHLRRGADGTAEITTTAATRFMKDLFTSTNGVGE 363
QY 60 AGR-----STLKDFIPVQGYAAGRDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
Db 364 IGRGIALTLFNLADTLGLPTELISAG-----GQLFYSRPVV-----SANGEPYVK 411
QY 110 IYVVOVEGEPDDASLAKLNGVTLNDGTLPGIERVNEPEWLPNRPPIRKRKSIPTS- 169
Db 412 L-YTSVENAQDQKGA-IPHDIDLGESRVVIQDDYDQHQDRPTSPAPSRPFSVLRAND 469
QY 169 --WLKITLYEGRNRQVRMTAHVGFPTLRIRYAMGSYTLDSLNG-----EWRDVT 218
Db 470 VLWLSLTAAYDQSTYGSSTGPGV-----YVSDSVTLNVNATGAQAVARSLDWTXVT 520

Search completed: June 10, 2000, 10:45:05
Job time: 6175 sec

THIS PAGE BLANK (USPTO)

A:Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: I64156
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-240 <TIGR>
A:Cross-references: GB:U32752; GB:IL42023; NID:g1573692; PIDN:AAC22354.1; PID:g1573697; T:Superfamily: conserved hypothetical protein HI1243

Query Match 51.3%; Score 598; DB 2; Length 240;
Best Local Similarity 59.2%; Pred. No. 5.3e-46;
Matches 119; Conservative 29; Mismatches 49; Indels 4; Gaps 2;

Qy	15	KTSFRKHVRVFSRQARTTPEQPTRVIRVFNKPDVLPQFTDEAGRSTLKDFIPQVG	74
Db	28	KSAYNFHRTFKPKLPKTSLSDE--TKVLENKPDVLTQFTDEGRATLKDFISIPV	84
Qy	75	YAAGRLRDSGLIVLTNDGVLAQLTQPKRTGKYYVQVEGDDASLAKLRNGVTIN	134
Db	85	YAAGRLRDSGLIITNNGELHRLADPFKTEKTYWQVEIPEETDLAQLRKGVELK	144
Qy	135	DGPTLPAGIRVNEPELWLPNRPPIRERKSIPTSWLKITLYEGRNQVRVMTAHVCFPTL	194
Db	145	DGVTKSAKVLISEPN-LWERNPIREKRNIPISWLEIKISEGRNQVRVMTAHIGFPTL	203
Qy	195	RLIRYANGSYTDSLANGWEWR	215
Db	204	RLVRVSMGLLSINGLENGSPR	224

RESULT 3
S74428
probable pseudouridylylate synthase slr0612 - *Synechocystis* sp. (PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S74428
R:kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

A:Reference number: S74322; MUID:97061201
A:Accession: S74428
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-185 <RAN>
A:Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAAL6580.1; PTD:d101731
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
A:Genetics:
A:Start codon: GTG

Query Match 33.4%; Score 389; DB 2; Length 185;
Best Local Similarity 52.8%; Pred. NO. 1.5e-27;
Matches 85; Conservative 16; Mismatches 42; Indels 18; Gaps 2;

QY	78	GRDRDSEGLVLVNDGVLOARLTQPGKRTGIYYVQVEGEPODDASLAKLRNGVTINDGP	137
Db		: : : : : : : : : : :	
	2	GRLDQDSEGLLLTSGKLQHRLAHEFAHQTYAQVEGSPDTDEPLEPLRGITFADYP	61
QY	138	TLPAGATERNNEEWLPRNPPTIRERKSISFTSWLKTLTEGRNRQVRMTAHVGFFTLRLI	197
Db		: : : : : : : : : :	
Db	62	TRPAIAKIITEPDFP-PRNPPPIRYRASIPTSWLSITLTEGRNRQVRMTAAVGFFTLRLV	120
QY	198	RX-----AMGSYTILDSLANGEWRDVTPKE	221
Db		: : : : :	
Db	121	RVQIQVIGRSPQOGKGKSAATWCLETGLELSPGOOWRPLTEWE	161

RESULT	4
S76104	

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 29-Sep-1999
C:Accession: S76104
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S76104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <RAN>
A:Cross-references: EMBL:DG3999; GB:AB001339; NID:g1001396; PIDN:BAAL0082.1; PID:dl01
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: conserved hypothetical protein HI1243

```

Query Match          25.6%; Score 298; DB 2; Length 249;
Best Local Similarity 38.5%; Pred. No. 2.8e-19;
Matches 72; Conservative 28; Mismatches 63; Indels 24; Gaps 3;

QY 44 ILFNKPYDVLVLPQFDEAGRSTLKDIFV-----QGVYAAGRLDSDGLVLVLTNDGVLOA 98
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 LLYNKPRDVLSTCDDPRGRVTVDLLPQDLQRGKGLPVGRGLDRNSTGAULTNDGELTL 126
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 99 RLTPQGRKRTGKIYYQVEGPPDASLAKLRNGVTLNDGPTLPAGIERVNPWEVLWRNP 158
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 RLTHPRYHLPKTYDVMLEGNPSDEDELEKWRSGMMDGKTLPATLEIVSENK ----- 179
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 159 IREKKSPTSLKTLTYEGNRQVRMTAVGPTPLRIYANGSYTLD-----LANGEW 214
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 -----DQHLVLTTEGRNQIRRLAEEGLTVLKLHRTTIGPQLHTRGKVLGSGQF 231
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 215 RDVTPKE 221
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 RFLSPAE 238
      : ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5
 G72400
 16S pseudouridylylate synthase - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 26-Aug-1999
 C:Accession: G72400
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, J.R.; Holt, J.E.; Marshall, M.; Sutton, G.; Tatus, K.L.; White, O.; Wilson, D.A.;
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.
 Nature 399, 323-329, 1999
 A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A:Reference number: A72200; MUID:99287316
 A:Accession: G72400
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-239 <ARN>
 A:Cross-references: GB:AE001708; GB:AE000512; NID:g4980740; PID:g4980760; TIGR:TM0364
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM0264
 C:Superfamily: conserved hypothetical protein H11243

```

Query Match      22.7%  Score 264.5;  DB 2;  Length 239;
Best Local Similarity 35.8%;  Prd. NO. 2.6e-16;
Matches 64;  Conservative 31;  Mismatches 65;  Indels 19;  Gaps 3;

44  ILFNKPYDLVPQFTDEAGRTLKDFIPVGQVYAAGLRDSEGLLVNDGSLVQLARLTQ 103
    ||| ||| : : : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
62  ILFKYQGVYITSTDKPHSETIMEFLPLKGIIFPVGLKDAEGLLITNDGDFAHRVIS 121
    ||| ||| : : : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

104  GKRTGKIYYVQGEPEDDASLAKLRNGVTINDGPTLPAGIERVNEPEWLNPANPIRERK 163
    ||| ||| : : : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```

Db 122 KWSVEKEIVVEGEVTEDEKLEKNGVTLRDG--FFAKARV-----E 163
QY 164 SIPTSWLKITLYEGNRQVRMTAHVGFPTLRLIRYAMGSYTL-DSLANGEWRDVTPE 221
Db 164 KLSNDTLKIVITEGYHOIKETRAVGLTKTRIGGLVLPDDMKPGYRFLSEEE 222
RESULT 6
D75462
ribosomal large subunit pseudouridine synthase B - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: D75462
R:White, O.; Eise, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250
A:Accession: D75462
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <WHI>
A:Cross-references: GB:AE001942; GB:AE000513; NID:g6458611; PIDN:AAF10472.1; PID:g645861
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0896
A:Map position: 1

Query Match 22.5%; Score 262; DB 2; Length 257;
Best Local Similarity 35.8%; Pred. No. 4.7e-16;
Matches 67; Conservative 23; Mismatches 73; Indels 24; Gaps 4;
QY 38 POPTRVILFNKPYDVLPOFTDEAGRSTLKDFIP-VQGVYAAGRLDRDSEGLVLTNDGVL 96
Db 61 PETVTYALY-KPGVFTTAHDEYGRNRVLDAMPDVGPHGVRLDKDSEGLVLTNDGDL 119
QY 97 QARLTQPKRGTGIYVQVEGE--PDDASLAKLRNGVTLDNGPTLPAGIERNVPEWLMP 154
Db 120 TLTLTTPRYGHEKAYRAWTEGREPTQAEVDLYVRGINDGPAQALSAAPEDGAY--- 177
QY 155 RNPPIREKSIPTSWLKITLYEGNRQVRMTAHVGFPTLRLIRYAMGSYTLDSLANGEW 214
Db 177 -----VVLGEGNRQVRRLMEALGHPVRLVRYVGLGLMDLNPGEY 219
QY 215 RDVTPE 221
Db 220 RELGPRD 226

RESULT 7
B70427
conserved hypothetical protein aq_1464 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 29-Sep-1999
C:Accession: B70427
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: B70427
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-249 <AQF>
A:Cross-references: GB:AE000741; NID:g2983841; PIDN:AAC07410.1; PID:g2983856; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_1464
C:Superfamily: conserved hypothetical protein H11243

Query Match 22.2%; Score 259; DB 2; Length 249;
Best Local Similarity 34.6%; Pred. No. 8.3e-16;
Matches 65; Conservative 28; Mismatches 77; Indels 18; Gaps 3;
QY 37 EPQPTRVILFNKPYDVLPOF-TDEAGRSTLKDFIP--VQGVYAAGRLDRDSEGLVLTND 93
Db 59 KPQRKRYIILNKPCYLTQGRSPDGRKTIIEELKIDIPERFVPVGRDLYNTEGLIITND 118
QY 94 GVLQARLTQPKRGTGIYVQVEGEPPDASLAKLRNGVTLDNGPTLPAGIERNVPEWLW 153
Db 119 GELANRLTLHRYKLPKYVLAIVEGKVDQKTLKRMKGIELEDGEAKPDNIRV----- 172
QY 154 PRNPPIREKSIPTSWLKITLYEGNRQVRMTAHVGFPTLRLIRYAMGSYTLDSLANGE 213
Db 172 -----RYEGKNTLEITFHEGRKHLVRFGLGAFGHKVRKRLARIAIGPIKLGKLSPGK 223
QY 214 RDVTPE 221
Db 224 WRELNGQE 231
RESULT 8
S45555
conserved hypothetical protein ypuL - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 29-Sep-1999
C:Accession: S45555; A69943
R:Sorokin, A.; Zumstein, E.; Azevedo, V.; Ehrlich, S.D.; Serror, P.
submitted to the EMBL Data Library, November 1993
A:Reference number: S45533
A:Accession: S45555
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-229 <SOR>
A:Cross-references: EMBL:L09228; NID:g410114; PIDN:AAA67493.1; PID:g410137
C:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
R.; Bron, S.; Broutillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: A69943
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-229 <KUN>
A:Cross-references: GB:G299116; GB:AL009136; NID:g2634723; PIDN:CAB14248.1; PID:ell855
A:Experimental source: strain 168
C:Genetics:
A:Gene: ypuL
C:Superfamily: conserved hypothetical protein H11243
Query Match 21.7%; Score 253; DB 2; Length 229;
Best Local Similarity 35.2%; Pred. No. 2.5e-15;
Matches 64; Conservative 31; Mismatches 69; Indels 18; Gaps 5;
QY 39 OPTRVILFNKPYDVLPOFTDEAGRSTLKDF---IPVQGVYAAGRLDRDSEGLVLTNDG 95
Db 59 EPVYFELY-KPRGVISAQDDKGRKVVTFDFKNIP-QRIYPIGRLDYDTSGLLLTNDGE 116
QY 96 LQARLTQPKRGTGIYVQVEGEPPDASLAKLRNGVTLDNGPTLPAGIERNVPEWLMP 155

Db 117 PANKLMPKPEIDITYVAKVGIGIPPKELLRLKRGIRLECKTAPAKAKLLS----- 169

QY 156 NPPIRERKSPTSLWKITLYEGRNRQVRMTAHVGFPTLRIRYAMGSYTLDLSLANGWR 215
||| || :||:||||||| | :|| | | :|
Db 169 ---LDKKKQ--TSIIQTIHEGRNQVRMFPAIGCHEVIKLKREYAFLNRLGLTGDA 223
QY 216 DV 217
Db 224 EL 225

RESULT 9
H64874
probable pseudouridylylase synthase ycil - Escherichia coli
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C:Accession: H64874
R.; Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H64874
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-291 <BLAT>
A:A:Cross-references: GB:AE000225; GB:U00096; NID:g1787523; PIDN:AAC74351.1; PID:g1787524;
A:Experimental source: Strain K-12, substrain MG1655
C:Genetics:
A:Gene: ycil
C:Superfamily: conserved hypothetical protein HI1243

Query Match 21.0%; Score 245; DB 2; Length 291;
Best Local Similarity 34.9%; Pred.No. 1.8e-14;
Matches 74; Conservative 23; Mismatches 79; Indels 36; Gaps 7;

QY 22 RVE-----RFSRQARTPEPQPTRVILFNKNPYDLVPQTDEAGRTLKDFIP-VQG 73
||| ||| :||:||||||| | :|| | | :|
Db 41 RVEVTGKIRIDGHILISRESAEQICRVLAYTKPEGELCTRNDPGEGRPTVFDRLPKLG 100
QY 74 V-YAAGRLDRDEGLLVLTNDGVLOARLTPQKRICKIYYVQVEGEPDDASIAKLNGV 131
||| ||| :||:||||||| | :|| | | :|
Db 101 ARNIANGRLDVNTCGLLFTFDGELANRLMHPGREVERAYRVFGQVDDAKLRDLSRGV 160
QY 132 TLNDGP----TLP-AGIERNEPEWLPNRPPIRERKS IPTSKIKTYEGRRNQVRMT 186
||| ||| :||:||||||| | :|| | | :|
Db 161 QLEDGPAAPTIFSGGEGINQ-----WYNVLTSGRRNREVRLW 200
QY 187 AHVGFPTLRIRYAMGSYTL-DSLANGEWDV 217
||| ||| :||:||||||| | :|| | | :|
Db 201 EAVGVQVSRLIRVRYGDPLPLPRGGWTTEL 232

RESULT 10
A64169
hypothetical protein HI1199 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: A64169
R.; Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
A.: Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.;
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: A64169
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-357 <TIGR>
A:Cross-references: GB:U032799; GB:L42023; NID:g1574122; PIDN:AAC22853.1; PID:g1574128;
A>Note: best homolog was a hypothetical protein from Buchnera aphidicola

OY 216 DVTPE 221
: : :
Db 221 PLTEEE 226

Search completed: June 10, 2000, 11:23:09
Job time: 8313 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 09:37:11 ; Search time 31.04 Seconds
(without alignments)
217.816 Million cell updates/sec

Title: US-09-252-691-7056

Perfect score: 1165

Sequence: 1 AIMRLIPENTWTKTSFK.....SYTLDLANGWRDVTPKEN 222

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	975	83.7	217	1 YMF6_ECOLI	P75966 escherichia
2	598	51.3	240	1 YMF6_HAEIN	P44827 haemophilus
3	474.5	40.7	261	1 Y612_SYNY3	P72581 synecocyst
4	298	25.6	249	1 Y361_SYNY3	O55578 synecocyst
5	259	22.2	249	1 Y664_AQUAE	O67444 aquifex aeo
6	253	21.7	229	1 RLUB_BACSU	P35159 bacillus su
7	245	21.0	291	1 YC1L_ECOLI	P37765 escherichia
8	244.5	21.0	357	1 YC1L_HAEIN	P45104 haemophilus
9	233.5	20.0	254	1 YH11_MYCTU	O33210 mycobacteri
10	217.5	18.7	256	1 YH11_MYCLE	O05658 mycobacteri
11	209.5	18.0	235	1 Y544_RICPR	Q92d06 rickettsia
12	206	17.7	231	1 RSUA_ECOLI	P33918 escherichia
13	201.5	17.3	290	1 YJBC_ECOLI	P32684 escherichia
14	200.5	17.2	238	1 Y554_AQUAE	O66829 aquifex aeo
15	198.5	17.0	157	1 Y7ZF_BACSU	O32058 bacillus su
16	186	16.0	232	1 RSUA_HAEIN	P45124 haemophilus
17	175.5	15.1	238	1 YC1L_BUCAP	P42395 buchnera ap
18	151	13.0	249	1 Y129_BORBU	O51155 borrelia bu
19	119.5	10.3	316	1 YH58_AQUAE	O67638 aquifex aeo
20	113.5	9.7	262	1 Y559_HELPY	P55986 helicobacte
21	111	9.5	253	1 YMDA_CHLAU	O45826 chloroflexu
22	97	8.3	326	1 RLUD_ECOLI	P73443 escherichia
23	96.5	8.3	327	1 YG2S_SYNY3	P74346 synecocyst
24	90.5	7.8	309	1 Y209_MYCPN	P75485 mycoplasma
25	88.5	7.6	302	1 YHCT_BACSU	P54604 bacillus su
26	88	7.6	366	1 QUEA_SYNY3	O55850 synecocyst
27	86.5	7.4	457	1 IREF_MOUSE	P70434 mus musculu
28	86.5	7.4	1275	1 RFEC_MYXXA	Q50864 myxococcus
29	85.5	7.3	591	1 RIB2_YEAST	Q12362 saccharomyc
30	85	7.3	376	1 YQCB_ERWCA	O47417 erwania car
31	85	7.3	404	1 YG3X_YEAST	P53294 saccharomyc
32	84	7.2	324	1 RLUD_HAEIN	P44445 haemophilus
33	83.5	7.2	2131	1 CINA_DROME	P35500 drosophila
34	83	7.1	485	1 VST2_HEVRH	Q00270 hepatitis e

ALIGNMENTS

RESULT 1

YMF6_ECOLI

ID YMF6_ECOLI STANDARD; PRT; 217 AA.

AC P75966;

DT 01-NOV-1997 (Rel. 35, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-FEB-2000 (Rel. 37, Last annotation update)

DE HYPOTHETICAL 24.9 KD PROTEIN IN TRMU-ICDA INTERGENIC REGION.

GN YMF6.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12/MG1655;

RX MEDLINE; 97426617.

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RT Science 277:1453-1474 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE; 97061202.

RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,

RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,

RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,

RA Yano M., Horiuchi T.;

RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:137-155 (1996).

CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.

CC STRONG, TO H.INFLUENZAE HI0694.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; A0000213; AAC74219.1; ALT_INIT.

DR EMBL; D90748; BAA35957.1;

DR EMBL; D90749; BAA35966.1;

DR ECOGENE; EGI3447; YMF6.

DR PROSITE; PS01149; PSI_RSU; 1.

KW Hypothetical protein.

SQ SEQUENCE 217 AA; 24880 MW; F7C7A7CEDC5FD3F6 CRC64;

Query Match 83.7%; Score 975; DB 1; Length 217;
Best Local Similarity 85.6%; Pred. No. 3.8e-79;

QY	3	MRQLITPENTWTKTSFKKHVRFSSRQATRRTPPEQPTRVILFNKPVDVLPQFTDEAGR	62
Db	1	MRQFIISNTMQKTSFNFHQVKFFSSQSRSTRKPENQPTRVILFNKPVDVLPQFTDEAGR	60
QY	63	STLKDFIPVQGVTAARLDRDSEGLLVLTNGVQLARLTQPGKRTCKIIYVQVVEGSPDDA	122
Db	61	KTLLKEFIPVQGVTAAGRLDRDSEGLVLVTNNGALQARLTQPGKRTCKIIYVQVVEGIPQD	120
QY	123	SLAKLRNGVTLNDGPITLPACIERVNEPEWLPWPNPIRERKSTPTSWLKITILEGNRQV	182
Db	121	ALSGALRNGVTLNDGPITLPAGAEVLDPWLPWPNPIRERKSTPTSWLKITILEGNRQV	180

```
QY      183  RRMTAHVGFPTLLRIYAMGSYTLDLANGWRDVT  218  
          ||| ||||| ||||| ||||| :||| ||||| :||  
Db      181  RRMTAHVGFPTLLRIYMGDYSLDLANGWEWFT  216
```

RESULT	2	YMF_C_HAEIN	STANDARD;	PRT;	240 AA.
ID	YMF_C_HAEIN				
AC	P44827;				
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	13-DEC-1998	(Rel. 37, Last annotation update)			
DE	HYPOTHETICAL PROTEIN HI0694.				

GN HI0694.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20;
RX MEDLINE: 95350630.

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodde A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Frichman J.L., Fuhrman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., SmallK.V., Fraser C.M., Smith H.O.,

RA	Venter J.C.;
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RT	

Science 269:496-512(1995).
-1- SIMILARITY BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC
CC
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcements> or send an email to license@sib-sib.ch).

CC
DR EMBL: U32752; AAC22354.1; -.
DR TIGR: HI0694; -.
DR PROSITE: PS01149; PSI_RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 240 AA; 27464 MW; 0AD99EC61F52C01C CRC64;

Query Match 51.3%; Score 598; DB 1; Length 240;
Best Local Similarity 59.2%; Pred. No. 7.8e-46;

[illegible][illegible]

RESULT 3
Y612_SINY3
ID Y612_SINY3
AC B72581.
STANDARD; PRT; 261 AA.

15-DEC-1998 (Rel. 37, Created)
15-DEC-1998 (Rel. 37, Last sequence update)
15-DEC-1998 (Rel. 37, Last annotation update)
HYPOTHETICAL 21.0 KD PROTEIN SLR0512.

OS *Synechocystis* sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.
[1]

RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Suglura M., Sasamoto S., Kinura T.,

RA Hosouchi T., Matsuno A., Murak A., Nakazaki N., Naruo K., Okumura S.
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synchocystis sp. strain PC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 RN [2]

RP CONCEPTUAL TRANSLATION.
RA Bairoch A.;
RL Unpublished observations (SEP-1998).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.

-!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS INTRODUCED IN POSITION 20 TO PRODUCE THIS ORF.

This SWISS-PROT entry is copyright. It is produced through a collaborator

CC CC CC CC CC
between the Swiss Institute of Bioinformatics and the EMBL outpost
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC CC CC CC CC CC
entitles requires a license agreement (See http://www.isd-sib.ch/anno
or send an email to license@isd-sib.ch).

DR EMBL; D90899; BAA16580.1; ALT_FRAME.

DR PROSITE; PS01149; PSI_RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 261 AA; 29561 MW; 601A453085C04A69 CRC64;

Query Match 40.7%; Score 474.5; DB 1; Length 261;
Best Local Similarity 50.5%; Pred. NO. 6.8e-35;
Matches 106; Conservative 23; Mismatches 56; Gaps
Indels 25;

[illegible]

QY	89 VLTNDGVQLQARLPQCKRTGKIYYVQVEGEDDASLAKLRNGVTLNDGPTLPAGTIEVNE	148
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
D6	89 LTTSGKLOHRLAHREFAHORTYFAOVESGPDEDI EPLRGITTAEDVTPRAIAKITFE	148

QY	149	PEWLPRNPPIREKSIPTSWLKITLYEGNRQVRMTAHVGFPTLRIRY	-----	200

Db	149	PDPP-PRNPPIYRASITPTSWLSITLTGSRNROVRMTAAVGFTPIRIRVVOITVGRSP		207

QY 200 -----AMGTYLDSLANGWDRVTPKE 221
 DB 208 QQKGKSAATWCTLEGLSPQWRPLTPWE 237

RESULT 4
 Y361_SYNY3 STANDARD; PRT; 249 AA.
 AC Q5578;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHETICAL 28.2 KD PROTEIN SLR0361.
 GN SLR0361.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96127529.
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N., Sugiura M., Tabata S.;
 RA "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome.";
 RL DNA Res. 2:153-166(1995).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
 CC EMBL; D63999; BAA10082.1; -
 DR PROSITE; PS01149; PSI_RSU; 1.
 DR PFAM; PF01479; S4; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 249 AA; 28228 MW; 315F208030180326 CRC64;

Query Match 25.6%; Score 298; DB 1; Length 249;
 Best Local Similarity 38.5%; Pred. No. 2.4e-19;
 Matches 72; Conservative 28; Mismatches 63; Indels 24; Gaps 3;

QY 44 ILFNKPYDVLFPQFTDEAGRSTLKDFIPV-----QGVYAAGRLDRDSEGLLVLTNDGVLQA 98
 DB 67 LLVKNPRDLVSTCDPRGKTKVLDLIPDLQGRGKGLHPVGRDLDRNSTGALLTNDGELTL 126

QY 99 RLTPGKRTGIYVQVEGEPODASLAKLRNGVTINDGPTLPAGIERVNEPEWLPNRP 158
 DB 127 RLTHPRYHLPTDYVWLGSENFDELEKWRSGMMLDGKTKLPATLEIVSENK----- 179

QY 159 IREKRSIPTSLKTLTYEGNRQVRMTAHVGFPTLRLIRYAMGSYTLDS-----LANGEW 214
 DB 179 -----DQIHLVTLTLEGRNRQIRRLAEELGLTVLKLHRTTIGPLQLHTRGKVLGSGQF 231

QY 215 RDVTPKE 221
 DB 232 RFLSPA 238

RESULT 5
 YE64_AQUAE STANDARD; PRT; 249 AA.
 AC O67444;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHETICAL PROTEIN AQ_1464.
 GN AQ_1464.

OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RC [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-VF5;
 RX MEDLINE; 98196666.
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RA "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus";
 RT Nature 392:353-358(1998).
 RL Nature 392:353-358(1998).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
 CC EMBL; AE00741; AAC07410.1; -
 DR PROSITE; PS01149; PSI_RSU; 1.
 DR PFAM; PF01479; S4; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 249 AA; 28518 MW; 29C7A2E7916FDE7C CRC64;

Query Match 22.2%; Score 259; DB 1; Length 249;
 Best Local Similarity 34.6%; Pred. No. 6.6e-16;
 Matches 65; Conservative 28; Mismatches 77; Indels 18; Gaps 3;

QY 37 EPQTRVLFNKPYDVLFPQFTDEAGRSTLKDFIPV-----VQGVYAAGRLDRDSEGLLVLTND 93
 DB 59 KPQRKRYILNKKPCYLTQLGRSPDKRTIELIKDIPERFVGRDLYNTEGLLITND 118

QY 94 GVLAQRTQPGKRTGIYVQVEGEPODASLAKLRNGVTINDGPTLPAGIERVNEPEWLM 153
 DB 119 GELANRILHPRYKLPKVVYALVEGKVDQKTLKRMKQGLEDEGFAKPDNIRIV----- 172

QY 154 PRNPPIREKRSIPTSLKTLTYEGNRQVRMTAHVGFPTLRLIRYAMGSYTLDSLANG 213
 DB 172 -----RYEGKNTLLEITTFHEGRKHLVRFGLGAFGHKVRKRLRIAGPIKGLKLSGPK 223

QY 214 WRDVTPE 221
 DB 224 WRELNGE 231

RESULT 6
 RLUB_BACSU STANDARD; PRT; 229 AA.
 AC P35159;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70)
 DE (PSEUDOURIDYLASE SYNTHASE) (URACIL HYDROLYASE).
 GN RLUB.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RC [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-168 / MABURG;
 RX MEDLINE; 95020538.
 RA Sotokin A.V., Zumstein E., Azevedo V., Ehrlich S.D., Serror P.;
 RA "The organization of the Bacillus subtilis 168 chromosome region between the spoVA and serA genetic loci, based on sequence data.";
 RL Mol. Microbiol. 10:385-395(1993).
 RN [2]
 RP CHARACTERIZATION.

DB 161 QLEDGPAARKTKFSGGEGINQ-----WYNTLTEGRNREVRRLW 200

QY 187 AHVGFPTLRLIRYAMGSYTL-DSLANGEWDV 217

DB 201 EAVGVQVSRILRVRYGDIPLKPLGRGGWTEL 232

RESULT 8

YCIL_HAEIN STANDARD; PRT; 357 AA.

AC P45104;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE HYPOTHETICAL PROTEIN H11199.

GN H11199.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RD / KW20;

RC MEDLINE: 95350630.

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Guyman J.V., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus

RT influenzae Rd.;"

RL Science 269:496-512(1995).

CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.

CC STRONG, TO E.COLI YCIL.

CC

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC

DR EMBL; U32799; AAC22853.1; -

DR TIGR: H11199; -

DR PROSITE; PS01149; PSI_RSU; 1.

DR PFWA; PF01479; S4; 1.

KW Hypothetical protein.

SQ SEQUENCE 357 AA; 40636 MW; 0A31F0D384CBBE82 CRC64;

Query Match

Best Local Similarity 21.0%; Score 244.5; DB 1; Length 357;

Matches 67; Conservative 24; Mismatches 70; Indels 19; Gaps 5;

QY 42 RVLEKNKPYDVLPOFTDEAGRTLKDFIP-VGVG--YAAGRLDRDSEGLLVLTNDGVLQA 98

DB 141 RVLMYKPGELCTRSDPGRATVDFRLPLTGSRIAVGRDLINTSGLLLTFTDDELAN 200

QY 99 RLTPQGRKTKIYYVQVEGPPDASLAKRLNGVTLNDGTLPLAGIERVNEPEWLPRNPP 158

DB 201 RLMHPSREVEREYSRVFGQVDDAMLARKVQLEDG-----LANFKEIKFTGGVG 252

QY 159 IREKSIPTSLWKITLYEGNRNVRMTAHVGPPTLRLIRYAMGSYTL-DSLANGEWDV 217

DB 253 INQ-----WYDVTLMGRNREVRRLWESQIQVSRILIRYRGNKLMKGLPRGGWEEM 305

RESULT 9

YH11_MYCTU

ID YH11_MYCTU STANDARD; PRT; 254 AA.

AC O33210;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DE HYPOTHETICAL 27.6 KD PROTEIN RV1711.

GN RV1711 OR MTC125.33.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RC MEDLINE: 98295987.

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jegeris K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,

RA Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence.;"

RL Nature 393:537-544(1998).

CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; Z98368; CAB10968.1; -

CC PROSITE; PS01149; PSI_RSU; 1.

CC PFWA; PF00849; YABO; 1.

CC PFWA; PF01479; S4; 1.

DR TUBERCULIST; RV1711; -

KW Hypothetical protein.

SQ SEQUENCE 254 AA; 27597 MW; EB2D95D7CF0DF26 CRC64;

Query Match 20.0%; Score 233.5; DB 1; Length 254;

Best Local Similarity 32.5%; Pred. No. 1.2e-13;

Matches 68; Conservative 25; Mismatches 81; Indels 35; Gaps 5;

QY 32 TRRTPEPQPTRV-----ILFNKPYDVLPOFTDEAGRTLKDFI--PVQG--- 74

DB 52 TRVDPQVAVRVVDGARVVLDDSLVTLANKPRGMHSTMSDDRGRPCIGDLIERKVRGTRK 111

QY 74 VYAAGRLDRDSEGLLVLTNDGVLQARLTOPGKRTGKIYYVQVEGPPDASLAKRLNGVTL 133

DB 112 LFHVGLDADTDELGLMLTNDGELAHLMHPSHEVPKTYLATVTGSPVPRGLGRTLAGIEL 171

QY 134 NPGPTLPAGIERVNEPEWLPRNPPPIRERKSIPTSLW-KITLYEGNRNVRMTAHVGP 192

DB 172 DDGPAFVDFAVVD-----AIPGKTLVRVTLHEGRNVRRLAANGPP 215

QY 193 TURLIRYAMGSYTLDSLANGEWDVTPKE 221

DB 216 VEALVRTDIGAVSLGKQRPQSVRLAKRSNE 244

RESULT 10

YH11_MYCLE

ID YH11_MYCLE STANDARD; PRT; 256 AA.

AC O05668; Q49886;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE HYPOTHETICAL 28.1 KD PROTEIN CB1351.03C.

GN MLCB1351.03C OR U0247G.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RA Murphy L., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 71-256 FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95117; CAB08276.1; -;
DR EMBL; U00021; NAA50926.1; -;
DR PROSITE; PS01149; PSI_RSU; 1.
DR PFAM; PF00849; YABO; 1.
DR PFAM; PF01479; S4; 1.
KW Hypothetical protein.
SQ SEQUENCE 236 AA; 28113 MW; 686746FA90FB695 CRC64;

Query Match 18.7%; Score 217.5; DB 1; Length 256;
Best Local Similarity 32.5%; Pred. No. 3.1e-12;
Matches 63; Conservative 25; Mismatches 71; Indels 35; Gaps 5;

QY 32 TRRTPEQPTVY-----ILFNKPYDVLQPTDEAGRSTLKDFTP--VQG--- 74
DB 54 TRVDVSVVRVGVKVVVDDSLVYALNKPGRMSTMSDDRGRCVGDLIERRVGNKK 113
QY 74 VYAGRLDRDSEGLVLTNDGVLQRLTPQKRTGKIYVOVEGEPDASLAKLRNGVTL 133
DB 114 LFHVGLRDADTGLLLNDGELARLHPHSHEVSKTYLATVGNVPRGLKKLSVGLLE 173
QY 134 NDGPTLPAGIERNPEWLPWRNPPIREKRSIPTSWL-KITLYEGRNQVRMTAHVGF 192
DB 174 DGGPAHVDDFAVD-----AIPGKTLVRLTLHEGRKRVRLTLTAAGFP 217
QY 193 TURLRYAMGSYTL 206
DB 218 VEMLVRTDIGAVSL 231

RESULT 11
ID Y544_RICPR STANDARD; PRT; 235 AA.
AC Q9ZD06;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN RP544.
GN RP544.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WADRID E;
RX MEDLINE; 99039499.
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."

RL Nature 396:133-140(1998).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235272; CAAL4993.1; -;
DR PROSITE; PS01149; PSI_RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 235 AA; 26815 MW; 9E09C0BAC46938A7 CRC64;

Query Match 18.0%; Score 209.5; DB 1; Length 235;
Best Local Similarity 28.3%; Pred. No. 1.4e-11;
Matches 53; Conservative 39; Mismatches 70; Indels 25; Gaps 5;

QY 39 QPTRVILENKPVDLPQFTDEAGRSTL-KDFIPVQGVYAAGRLDRDSEGLVLTNDGVILQ 97
DB 59 QKPRLWIYKPVGLITTHKDLPLSRKTVFQQLICLPVIGRLDLNSEGILLTNSGLA 118
QY 98 ARLTQPKRTGKIYVOVEGEPDASLAKLRN---GVTLNDGPTLPAGIERNPEWLPW 153
DB 119 HQFEMPASKLRKRVNVRAYGNPNILLKNYKNIKIDGIFYN-----PHSIKL 166
QY 154 PRNPPIREKRSIPTSWLKITLYEGRNQVRMTAHVGFVTLRLIRYAMGSYTLSDSLANGE 213
DB 166 -----LFQNKNS--NSWFVYLFEGKNREIRIFEFGLQVKNKILRIQYGALKIGNLKPGN 218
QY 214 WRDVTPK 220
DB 219 YQEISNK 225

RESULT 12
ID RSUA_ECOLI STANDARD; PRT; 231 AA.
AC P33918;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RIBOSOMAL SMALL SUBUNIT PSEUDOURIDINE SYNTHASE A (EC 4.2.1.70) (16S
DE PSEUDOURIDYLATE 516 SYNTHASE) (16S PSEUDOURIDINE 516 SYNTHASE) (URACIL
DE HYDROLYASE).
GN RSUA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-13, AND CHARACTERIZATION.
RX MEDLINE; 95337112.
RA Wziesinski J., Bakin A., Nurse K., Lane B.G., Ofengand J.;
RT "Purification, cloning, and properties of the 16S RNA pseudouridine
RT 516 synthase from Escherichia coli."
RL Biochemistry 34:8904-8913(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

Job time: 6514 sec

Best Local Similarity 27.4%; Pred. No. 9e-11; Matches 58; Conservative 43; Mismatches 76; Indels 35; Gaps 5;

QY 20 KHRVERFSSRQATRTPEPQPTTRVILFNKPYDVL-----POFTDEAGRSLKDFIP 70
Db 38 EYRVKEGEVEVEGSKVPKKNVLMYLPKGYLSTTEEDAKYPSFLE-----LIREHFP 92
QY 71 VOGVYAAGRLDRDSGLLVLTNDGVLQARLTQPGKRTKIYVQVEGEPDDASLAKLRNG 130
Db 93 SRKLSFAGRLDVAEGLLITDDGELAHRLTHPKWKEKEYIVRLDRDIGDEELKKLYE- 152
QY 131 VTINDGPTLPAGIERVNEPEWLPNRPPIRERKSIPTSWLKITLYEGRNQVRRMTAHVG 190
Db 152 VKLEEXPVQVIAEKLS-----GDTVKAILTEGRHHVVKRLFKAVG 192
QY 191 FPTLRIRYAMGSYTLTLD-SLANGENRDTVPKE 221
Db 193 HNVVYLKRTRVGNLRLDENNEPCWRELTEEE 224

RESULT 15
YTZF_BACSU STANDARD; PRT: 157 AA.
AC O32068;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 17.7 KD PROTEIN IN AMYX-OPUD INTERGENIC REGION.
GN YTZF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 299119; CAB14981.1;
DR SUBTILIS; BG13940; YTZF.
DR PROSITE; PS01149; PSI_RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 157 AA; 17695 MW; D378BFA41E89DCE1 CRC64;

Query Match 17.0%; Score 198.5; DB 1; Length 157;
Best Local Similarity 36.2%; Pred. No. 7.9e-11;
Matches 54; Conservative 25; Mismatches 53; Indels 17; Gaps 4;
QY 75 YAAGRLDRDSGLLVLTNDGVLQARLTQPGKRTKIYVQVEGEPDDASLAKLRNGVTLN 134
Db 15 FPAGRLDKDTGFLLLTNDGGLAHRLSPKHKVPTVEVHLKSOISREDISDLGTGYIE 74
QY 135 DG-PTLPAGIERVNEPEWLPNRPPIRERKSIPTSWLKITLYEGRNQVRRMTAHVGFTT 193
Db 75 GGYTKPAKAE-----IKTNSGNTV-IYLTITEGKHQVQMAKAVGNEV 119
QY 194 LRLIRYAMGSYTLTLD-SLANGENRDTVPKE 221
Db 120 VYLRKLSMGRVSLDPALAPGEYRELTEEE 148

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 09:36:29 ; Search time 53.1 Seconds
(without alignments)
289.871 Million cell updates/sec

Title: US-09-252-691-7056

Perfect score: 1165

Sequence: 1 AMRLQIPMTMTKTSFRK.....SYTLDSLANGEWDVTPKEN 222

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 59334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTEMBL12:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264.5	22.7	239	2 Q9WVA2	Q9WVA2 thermotoga
2	209.5	18.0	235	2 Q9ZD06	Q9ZD06 rickettsia
3	178.5	15.3	235	2 Q9Z742	Q9Z742 chlamydia p
4	176	15.1	338	5 Q96270	Q96270 plasmodium
5	153.5	13.2	241	2 Q84728	Q84728 chlamydia t
6	120	10.3	261	2 Q83472	Q83472 treponema p
7	112.5	9.7	262	2 Q9ZJG0	Q9ZJG0 helicobacte
8	112.5	9.7	304	2 Q9XJ38	Q9XJ38 thermotoga
9	107.5	9.2	1464	5 Q9XZ15	Q9XZ15 leishmania
10	95.5	8.2	217	2 Q50765	Q50765 mycobacteri
11	88.5	7.6	454	1 Q54664	Q54664 halobacteri
12	88	7.6	303	2 Q84108	Q84108 chlamydia t
13	86.5	7.4	454	5 Q9XZ07	Q9XZ07 caenorhabdi
14	84.5	7.3	473	2 Q9Z664	Q9Z664 zymomonas m
15	84.5	7.3	660	5 Q24717	Q24717 drosophila
16	84.5	7.3	689	5 Q24715	Q24715 drosophila
17	84.5	7.3	693	5 Q24712	Q24712 drosophila
18	84.5	7.3	701	5 Q24713	Q24713 drosophila
19	84.5	7.3	704	5 Q24716	Q24716 drosophila
20	84.5	7.3	714	5 Q24714	Q24714 drosophila

21	84.5	7.3	946	10 Q42571	Q42571 arabidopsis
22	84.5	7.3	946	10 Q42587	Q42587 arabidopsis
23	84.5	7.3	1020	10 Q37145	Q37145 arabidopsis
24	84.5	7.3	1020	10 Q37146	Q37146 arabidopsis
25	84	7.2	660	12 Q68985	Q68985 hepatitis e
26	84	7.2	2354	5 Q24135	Q24135 drosophila
27	83.5	7.2	2077	5 Q24528	Q24528 drosophila
28	83.5	7.2	2110	5 Q24532	Q24532 drosophila
29	83.5	7.2	2118	5 Q24531	Q24531 drosophila
30	83.5	7.2	2121	5 Q24529	Q24529 drosophila
31	83	7.1	660	12 Q81878	Q81878 hepatitis e
32	82.5	7.1	4151	2 Q53490	Q53490 mycobacteri
33	82	7.0	235	2 Q49848	Q49848 mycobacteri
34	81.5	7.0	1091	13 Q9RH41	Q9RH41 gallus gall
35	81.5	7.0	4162	13 Q89918	Q89918 gallus gall
36	81	7.0	303	10 Q80967	Q80967 arabidopsi
37	81	7.0	419	1 Q26396	Q26396 methanobact
38	81	7.0	1129	4 Q60292	Q60292 homo sapien
39	80	6.9	86	10 Q80966	Q80966 arabidopsi
40	80	6.9	436	12 Q9WI48	Q9WI48 hepatitis e
41	80	6.9	660	12 Q59419	Q59419 hepatitis e
42	80	6.9	1145	10 Q04146	Q04146 acetabulari
43	79.5	6.8	262	5 Q16582	Q16582 caenorhabdi
44	79	6.8	660	12 Q69411	Q69411 hepatitis e
45	79	6.8	660	12 Q81871	Q81871 hepatitis e

ALIGNMENTS

RESULT 1

Q9WVA2 ID Q9WVA2 PRELIMINARY; PRT; 239 AA.
AC Q9WVA2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 16S PSEUDOURIDYLATE SYNTHASE.
GN TM0264.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99287316.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA McDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RT "Evidence for lateral gene transfer between Archaea and bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA McDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE001708; AAC35352.1; -
SQ SEQUENCE 239 AA; 27160 MW; 511D06DD CRC32;

Query Match 22.7%; Score 264.5; DB 2; Length 239;

Best Local Similarity 35.8%; Pred. NO. 4.3e-16;

Matches 64; Conservative 31; Mismatches 65; Indels 19; Gaps 3;

QY 44 ILFNKPYDVLPPQFDTAGRSTLKDIFPQGVYVYAGRLDRSEGLVLTNDGVLQARLTOP 103

DB 62 ILFYKPSGYTSTKDPHSETIMEFLPKGIFPVGRDKDAEGLLIITNDGFAHRVISP 121

```
QY 104 GKRTGKIYVQVEGEPDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLPNPPPIRERK 163
DB 122 KWSVEKIYVQVEGTEDEKLEKNGVTLRDG--FFAKARV-----E 163
QY 164 SIPTSWLKITLYEGRNQRVRRMTAHVGPTRLIRYAMGSYTL-DSLANGEWRDVTPE 221
DB 164 KLSNDTLAIVTEGKYHOKRWTRAAGVGLKTVHLKTRIGGLVLPDDMKPGGEYRLSEEE 222

RESULT 2
Q92D06 PRELIMINARY; PRT; 235 AA.
AC Q92D06;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 26.8 KD PROTEIN.
GN RP544.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99039499.
RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA SICHERTITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA ERIKSSON A.S., WINKLER H.H., KORLAND C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA ANDERSSON S.G.E.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ235272; CAA14993.1;
DR PROSITE; PS01149; PSI RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 235 AA; 26815 MW; E5704F25 CRC32;
```

```
Query Match 18.0%; Score 209.5; DB 2; Length 235;
Best Local Similarity 28.3%; Pred. No. 3.2e-11;
Matches 53; Conservative 39; Mismatches 70; Indels 25; Gaps 5;

QY 39 QPTRVILFNKPYDVLPOFTDEAGRSTL-KDFIPVQGVYAAGRLDRDSEGLLVLTNDGVILQ 97
DB 59 QKPLRIYKPVGLITHTKDP+SRKTVFQGLIGLPRVISIGRLDNLSEGLLLTNSGDLA 118

QY 98 ARLTQPKGRTGKIYVQVEGEPDASLAKLRN---GVTNDGPTLPAGIERVNEPEWLW 153
DB 119 HQFEMPASKLRYNVRAGNPNILKNYKNLKIDGIFYN-----PHSIKL----- 166

QY 154:PRNPPPIRERKSPTSWLKITLYEGRNQRVRRMTAHVGPTRLIRYAMGSYTLDSLANGE 213
DB 166 -----LRONKS--NSWFVVLFEKGRNRIIRIFVEYGLQVNLRIQYGAALKIGNLKPGN 218

QY 214 WRDVTPE 220
DB 219 YQEISNK 225

RESULT 3
Q92742 PRELIMINARY; PRT; 235 AA.
AC Q92742;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE PREDICTED PSEUDOURIDINE SYNTHASE.
GN YJBC.
```

```
Query Match 15.1%; Score 176; DB 5; Length 338;
Best Local Similarity 27.1%; Pred. No. 5e-08;
Matches 57; Conservative 38; Mismatches 83; Indels 32; Gaps 6;

QY 44 ILFNKPYDVLPOFTDEAGRSTLKDFFIPVQ-----GVYAAGRLDRDSEGLLVLTNDGVILQ 98
DB 120 IYVLPKGLLCTSDENKRNKSIYTLFPEEMLOKRYLTVVGRDRLNTSGVLLLTNDYAWVN 179

QY 99 RLTOPGKRTGKIYVQVEGEPDASLAKLRNGVTL-NDGPTLPAGIERVNE----- 149
DB 180 KLTHPKYQRTYRVHIEGVPKMNALRELARGIYLEEDKTPKKIYKNSREKSNIDD 239

QY 149 -----PWLWPNPP-----TREK-----SPTSWLKITLYEGRNQRVRRMTAHVGP 192
DB 240 KKKKMSKMKKKTNPFAFTEILREKIKEDTKTKITVLNISIKEGRNQIRKMQQINQP 299
```

```
OS Chlamydia pneumoniae.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029.
RA KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,
RA GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;
RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis."
RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE001667; AAD19002.1;
DR PROSITE; PS01149; PSI RSU; 1.
SQ SEQUENCE 235 AA; 25942 MW; 80240286 CRC32;
```

```
Query Match 15.3%; Score 178.5; DB 2; Length 235;
Best Local Similarity 27.0%; Pred. No. 1.8e-08;
Matches 40; Conservative 26; Mismatches 63; Indels 19; Gaps 1;

QY 74 VYAAAGRLDRDSEGLLVLTNDGVILQARLTQPKGRTGKIYVQVEGEPDASLAKLRNGVTL 133
DB 98 VFTVGRLDKETSGLLVLTNDGEFANKIHPSSGITREYLLKVRSDYSAKDLGLKMEGTFI 157

QY 134 NDGPTLPAGIERVNEPEWLPNPPPIRERKSPTSWLKITLYEGRNQRVRRMTAHVGPFT 193
DB 158 DGKRVPSVTKIRR-----GTVKIVSEGKKEIRLFADAAGFFI 198

QY 194 LRLIRYAMGSYTLDSLANGEWRDVTPE 221
DB 199 LELKRIRIGSLVGLGLEYGEYRELTDAAE 226

RESULT 4
Q96270 PRELIMINARY; PRT; 338 AA.
AC Q96270;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE PSEUDOURIDINE SYNTHETASE (RSUA FAM.), 1ST EUR. MEMBER (OO).
GN PFE0890C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99021743.
RA GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
RA KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEDERSON J.,
RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
DR EMBL; AE001423; AAC71967.1;
SQ SEQUENCE 338 AA; 40297 MW; AFDC23C5 CRC32;
```

```
Query Match 15.1%; Score 176; DB 5; Length 338;
Best Local Similarity 27.1%; Pred. No. 5e-08;
Matches 57; Conservative 38; Mismatches 83; Indels 32; Gaps 6;

QY 44 ILFNKPYDVLPOFTDEAGRSTLKDFFIPVQ-----GVYAAGRLDRDSEGLLVLTNDGVILQ 98
DB 120 IYVLPKGLLCTSDENKRNKSIYTLFPEEMLOKRYLTVVGRDRLNTSGVLLLTNDYAWVN 179

QY 99 RLTOPGKRTGKIYVQVEGEPDASLAKLRNGVTL-NDGPTLPAGIERVNE----- 149
DB 180 KLTHPKYQRTYRVHIEGVPKMNALRELARGIYLEEDKTPKKIYKNSREKSNIDD 239

QY 149 -----PWLWPNPP-----TREK-----SPTSWLKITLYEGRNQRVRRMTAHVGP 192
DB 240 KKKKMSKMKKKTNPFAFTEILREKIKEDTKTKITVLNISIKEGRNQIRKMQQINQP 299
```


QY /
9 PENTWKTSPRKH-----RVRFSS-----RQATRRTPQP-----TRVILENKP 49
 ! ! . : | : ||| | : | : | : :||| : :|||

RESULT	14	
Q9Z664		
ID	Q9Z664	PRELIMINARY; PRT; 473 AA.
AC	Q9Z664;	
DT	01-MAY-1999	(TrEMBLrel. 10, Created)
DT	01-MAY-1999	(TrEMBLrel. 10, Last sequence update)
DT	01-MAY-1999	(TrEMBLrel. 10, Last annotation update)
DE	HYPOTHETICAL 52.7 KD PROTEIN.	
OS	Zymomonas mobilis.	
OC	Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;	

WS3871
ID W53871 standard; peptide; 22 AA.
AC W53871;
DT 13-JUL-1998 (first entry)
DE Gravin polypeptide fragment, residues 295-316.
KW Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW CAMP-dependent protein kinase; protein kinase C; autolysosomal disease;
KW Myasthenia gravis; nicotinic acetylcholine receptor.
OS Homo sapiens.
PN US5741890-A.
PD 21-APR-1998.
PF 19-DEC-1996; 769309.
PF 19-DEC-1996; US-769309.
(UYOR-) UNIV OREGON HEALTH SCI.
KLAUCK TM, Nauer JB, Scott JD;
WPI: 98-26052/23.
DR New polypeptide fragments of protein kinase binding protein gravin -
PT are useful for the study of modulation of action between gravin and
PT protein kinase(s)
PS Example 3; Column 10; 32pp; English.
CC This sequence corresponds to residues 295-316 of gravin, and represents
CC a polypeptide of the invention. The polypeptides are fragments capable of
CC binding to type II regulatory subunit of cAMP-dependent protein kinase
CC (PKA). Gravin is a kinase anchoring protein that binds to type II
CC regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC develops antibodies against their own nicotinic acetylcholine receptors.
CC The polypeptides are useful for providing analogues of gravin in the
CC study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC interactions between gravin and kinase. The peptides are involved in the
CC modulation of gravin-kinase interactions.
SQ Sequence 22 AA;

Query Match 2.7%; Score 6; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 KYSFRK 20
|||||
15 KYSFRK 20

RESULT 10
WI0941
ID WI0941 standard; peptide; 35 AA.
AC WI0941;
DT 10-NOV-1997 (first entry)
DE Polyclonal anti-ferritin binder sequence, C28, from R26 library.
KW Functional surrogate; analyte; affinity receptor; immunoreactive group;
KW mimic; homogeneous immunoassay; detection; diagnostic analyte; Chlamydia;
KW cardiac marker; tumour marker; allergen; hormone; fertility; myoglobin;
KW pregnancy; infectious disease; ferritin; myosin light chain; troponin;
KW follicle stimulating hormone; human; growth hormone; immunoglobulin E;
KW prolactin; parathyroid hormone; placental lactogen; hepatitis antigen;
KW antibody; chorionic gonadotropin; luteinising hormone; cytomagalovirus;
KW Streptococcus; rubella; toxoplasma; DK-MB; prostate-specific antigen;
KW carcinoembryonic antigen; alpha-fetoprotein; herpes virus; Cal25.
OS Synthetic.
PN WO9641172-A1.
PD 19-DEC-1996.
PF 07-JUN-1996; U10498.
PF 07-JUN-1996; US-476375.
PA (CYTO-) CYTOGEN CORP.
PI Carter JM, Lee-Own FV;
DR WPI: 97-077284/07.
DR N-PSDB; T48787.
PT Labelled functional surrogate of an analyte - useful as competitor
PT molecule in affinity assays, esp. for detecting large macromolecules
PT such as ferritin
PS Disclosure; Page 118/2; 156pp; English.
CC This sequence represents a polyclonal anti-ferritin binder sequence from
CC the R26 library (C series) which may be used in the conjugate of the

CC invention. The novel labelled conjugate comprises at least one label
CC attached to a functional surrogate of an analyte of interest. The
CC surrogate is capable of competing effectively with the analyte for a
CC limiting amount of an affinity receptor for the analyte. The conjugate
CC exhibits an activity that is altered upon interaction with the affinity
CC receptor and this activity can be measured and related to the amount of
CC the analyte present in a sample. Functional surrogates such as this have
CC an immunoreactive group that allows the surrogate to compete effectively
CC and with the analyte for a limiting amount of its affinity receptor.
CC Functional surrogates are able to mimic naturally occurring analytes.
CC They can be labelled for use in standard competitive affinity assays
CC (esp. homogeneous immunoassays) for detecting large macromolecules such
CC as polypeptides, polysaccharides, polynucleotides, glycoproteins and
CC lipid-containing macromolecules, as well as small haptens. Typical
CC diagnostic analytes for detection include cardiac or tumour markers,
CC allergens, hormones related to fertility-pregnancy or analyses associated
CC with infectious disease. In particular, the assays are useful for
CC detecting ferritin, follicle stimulating hormone, human growth hormone,
CC immunoglobulin E, prolactin, parathyroid hormone, human placental
CC lactogen, hepatitis antigens or antibodies against them, human
CC chorionic gonadotropin, human luteinising hormone, cytomagalovirus,
CC Chlamydia, Streptococcus a, rubella, toxoplasma, herpes virus, DK-MB,
CC myoglobin, myosin light chain, troponin, carcinoembryonic antigen,
CC alpha-fetoprotein, prostate-specific antigen and Cal25 (a tumour marker).
SQ Sequence 35 AA;

Query Match 2.7%; Score 6; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 AAGRLD 81
|||||
Db 17 AAGRLD 22

RESULT 11
WS9222
ID WS9222 standard; peptide; 39 AA.
AC WS9222;
DT 27-AUG-1998 (first entry)
DE Mouse gamma interferon peptide fragment muIFN-gamma (95-133)S.
KW Interferon-gamma; IFN-gamma; mouse; agonist; intracellular activity;
KW inducer; macrophage; antiviral; major histocompatibility complex;
KW MHC class II; treatment; disease.
OS Mus sp.
PN US5770191-A.
PD 23-JUN-1998.
PF 24-MAY-1995; 449639.
PF 24-MAY-1995; US-449639.
PA (UYEL) UNIV FLORIDA.
PI Johnson HM, Szente BE;
DR WPI: 98-376742/32.
DR Interferon-gamma agonists with antiviral activity - comprising
PT C-terminal fragment of interferon-gamma
PS Example 1; Column 3-4; 13pp; English.
CC This sequence is a peptide fragment of mouse gamma interferon (IFN-gamma)
CC which corresponds to amino acids 95 to 133 of the full length protein and
CC is a scrambled version of the peptide represented in W59221. This peptide
CC is used in a method which identifies fragments which, once internalised
CC into a cell act as agonists and have biological activity which is
CC comparable to the full length IFN-gamma protein. IFN-gamma agonists can
CC bind to the cytoplasmic domain of the IFN-gamma receptor and exhibit
CC intracellular interferon-gamma activity, inducing macrophage antiviral
CC activity and inducing major histocompatibility complex (MHC) class II
CC expression. They can be used to treat a variety of clinically relevant
CC disease states in animals and humans.
SQ Sequence 39 AA;

Query Match 2.7%; Score 6; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 KHRVER 25
Db 24 KHRVER 29

RESULT 12

ID R81962 standard; Protein; 60 AA.
AC R81962;
DE ORF encoded by exon 9 of rice pyruvate orthophosphate dikinase gene.
KW rice; promoter; exon; tissue-specific; season-specific; PPKK;
OS pyruvate orthophosphate dikinase; C4 photosynthesis; C3 plant.
PN Oriza sativa L.
PD 707184657-A.
PF 25-JUL-1995.
PR 28-DEC-1993; 335671.
PA (NITA-) JAPAN TAFU GURASU KK.
DR WPI: 95-287968/38.
DR N-PSDB; Q98249.
PT Pyruvate orthophosphate dikinase (C4 photosynthesis-related) gene
expression.
PT and promoter in rice - shows tissue-specific and seasonal specific
expression.
Claim 2; Page 9; 26pp; Japanese.
The sequence is the open reading frame encoded by exon 9 of rice pyruvate
orthophosphate dikinase (PPDK) gene, and which encodes precursor peptide.
The gene shows tissue-specific (in rice leaf and rice arista) and
seasonal specific expression. The PPKK gene may give a C4 photosynthesis
function to a C3 plant.
SQ Sequence 60 AA;

Query Match

Best Local Similarity 2.7%; Score 6; DB 1; Length 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 GKRTGK 109
Db 19 GKRTGK 24

RESULT 13

ID W27822 standard; Protein; 71 AA.
AC W27822;
DE 21-JUL-1998 (first entry)
KW Staphylococcus aureus protein of unknown function.
KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome.
PN Staphylococcus aureus.
PD W09730070-AI.
PF 21-AUG-1997.
PR 19-FEB-1997; U02318.
PR 20-FEB-1996; US-011888.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Black Mt, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR WPI: 97-424969/39.
DR N-PSDB; T83787.
PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
to isolate antimicrobial compounds, and in vaccines against S.
PT aureus infection
PS Claim 6; Page 294; 989pp; English.
CC The present sequence represents a Staphylococcus aureus protein of
unknown function. The DNA sequence was isolated from a library of
clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
be used in the construction of ribozymes and antisense sequences to
control the expression of Staphylococcal genes. The DNA sequence is
also useful as a source of regulatory elements for the control of

CC bacterial gene expression. The present protein may be used to produce
CC vaccines to enable a host to produce specific antibodies with
CC antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.
SQ Sequence 71 AA;

Query Match 2.7%; Score 6; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 KLRNGV 131
Db 36 KLRNGV 41

RESULT 14

ID W59026 standard; Protein; 88 AA.
AC W59026;
DE 14-SEP-1998 (first entry)
DE Colicin V.
KW Colicin V; bacteriocin; signal peptide; protein secretion;
KW preservative; food spoilage; lactic acid bacterium; antibacterial.
OS Escherichia coli.
FH Key
FT Disulfide bond 76..87
PN W09809639-AI.
PD 12-MAR-1998.
PF 05-SEP-1997; U15609.
PR 05-SEP-1996; US-026257.
PA (UYAL-) UNIV ALBERTA.
PI Franz CM, Greer GG, Leisner JJ, McCormick JK, McMullen LM,
PI Pkns J, Poona, Stiles ME, Van Belkum MJ, Vederas JC,
PI Worobo RJ, Worobo RW;
DR WPI: 98-193319/17.

PT Bacterial growth inhibiting peptide(s) enterocin 900 or brochocin C
- used to inhibit growth of susceptible bacteria in e.g. foodstuff,
PT living animal, food preparation area or fermentation vessel
PS Example 3; Page 140; 174pp; English.
CC This polypeptide comprises colicin V, a bacteriocin produced by
CC Escherichia coli. Colicin V secretion in Lactococcus lactis hosts
CC was obtained by fusion to the leucocin A leader peptide (see
CC W59013). The invention provides secretion vectors in which the
CC structural and immunity genes of bacteriocins, such as
CC carnobacteriocin B2 (see W59029-31), colicin V, leucocin A and
CC brochocin-C (see W59021), are fused to a signal peptide sequence,
CC such as the colicin V signal peptide, allowing production and
CC secretion of the bacteriocin(s) by host and heterologous bacteria.
CC Bacteriocin cassettes encoding two or more bacteriocins are used to
CC tailor the antibacterial spectrum of a producer strain to target a
CC range of spoilage or pathogenic bacteria, e.g. E. coli and
CC Salmonella. Using the technology of multiple bacteriocin
CC production and delivery using lactic acid bacteria, a range of
CC bacteriocins can be produced in situ in human food, animal feed,
CC the gastrointestinal tract of humans and animals, and in the female
CC genital tract.
SQ Sequence 88 AA;

Query Match 2.7%; Score 6; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 YAAGRL 80
Db 70 YAAGRL 75

RESULT 15

W94884

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	7	3.2	192	1	W55509	H. pylori ORF 04ge	
2	7	3.2	208	1	W55364	H. pylori ORF 04ge	
3	7	3.2	241	1	W87996	A human MCG18 prot	
4	7	3.2	339	1	W71482	Helicobacter polyph	
5	7	3.2	352	1	W23307	Lipase modulator f	
6	7	3.2	398	1	W55695	H. pylori ORF 13ae	
7	7	3.2	416	1	R4434	Subtilisin-like se	
8	7	3.2	416	1	R4434	Subtilisin-like se	
9	6	2.7	764	22	W53871	Melon aconitase en	
10	6	2.7	764	22	W53871	Gravin polypeptide	
11	6	2.7	35	1	W10941	Polyclonal anti-fe	
12	6	2.7	39	1	W52222	Mouse gamma interf	
13	6	2.7	60	1	R81962	ORF encoded by exo	
14	6	2.7	71	1	W27822	Staphylococcus aur	
15	6	2.7	88	1	W59026	Colicin V. Bacteri	
16	6	2.7	88	1	W94884	Colicin V cleaved	
17	6	2.7	94	1	R74035	Synthetic D32.39 e	
18	6	2.7	103	1	W59025	Colicin V prepepti	
19	6	2.7	103	1	W94883	Colicin V prepepti	
20	6	2.7	127	1	W47279	Papillomavirus maj	
21	6	2.7	128	1	W47280	Papillomavirus maj	
22	6	2.7	128	1	W47281	Papillomavirus maj	
23	6	2.7	132	1	W07861	(DSM 10103) human	
24	6	2.7	136	1	W07854	(DSM 10097) human	
25	6	2.7	165	1	W07636	P. suis leukotoxin	
26	6	2.7	171	1	W31709	Alternate transcri	
27	6	2.7	211	1	W29676	X25 gene (exons 1-	
28	6	2.7	211	1	W29676	Homo sapiens clone	
29	6	2.7	226	1	W21019	H. pylori cell env	
30	6	2.7	242	1	W31730	Mycobacterium tube	
31	6	2.7	264	1	W77717	3-Isopropylmalate	
32	6	2.7	292	1	W53862	Gravin polypeptide	
33	6	2.7	321	1	W72133	HSV-2 strain SB5 C	
34	6	2.7	321	1	W72183	HSV-2 strain SB5 C	

545	5	2.3	114	1	Y06831	Mouse MAb 6A4 muta	618	5	2.3	138	1	R42211	CRABP-II gene prod
546	5	2.3	115	1	P90149	Sequence of hepati	619	5	2.3	138	1	R68543	Hepatitis C virus
547	5	2.3	116	1	P92032	Sequence encoded i	620	5	2.3	138	1	W00771	C-terminally trunc
548	5	2.3	117	1	P50102	Atrial natriuretic	621	5	2.3	138	1	W07434	Large monomer of m
549	5	2.3	118	1	W02149	Anti-obesity prote	622	5	2.3	138	1	W00606	Large monomer of m
550	5	2.3	119	1	W26543	Trypanosoma cruzi	623	5	2.3	138	1	W01689	HCV protease activ
551	5	2.3	120	1	W00616	Large monomer of m	624	5	2.3	138	1	W14774	HCV protease activ
552	5	2.3	121	1	R92732	Obesity protein C-	625	5	2.3	138	1	W14775	HCV protease activ
553	5	2.3	122	1	W00067	C-terminally trunc	626	5	2.3	138	1	W46393	Amino acid sequenc
554	5	2.3	123	1	W02150	Anti-obesity prote	627	5	2.3	138	1	W42425	Pyruvate dehydroge
555	5	2.3	124	1	W02139	Anti-obesity prote	628	5	2.3	138	1	W65025	Thuja plicata diri
556	5	2.3	125	1	W77551	Nickel-binding per	629	5	2.3	138	1	W97605	Amino acid sequenc
557	5	2.3	126	1	Y11895	Human 5' EST seque	630	5	2.3	138	1	W04922	Mycobacterium spec
558	5	2.3	127	1	W03946	DNA fragment vk65..	631	5	2.3	139	1	W25686	E. coli beta-galac
559	5	2.3	128	1	Y13025	Human secreted pro	632	5	2.3	139	1	W70271	Brugia malayi tran
560	5	2.3	129	1	R13161	Pregnancy-specific	633	5	2.3	139	1	W76342	Papillomavirus mai
561	5	2.3	130	1	R14539	Hepatitis C Virus	634	5	2.3	139	1	W98012	Hypersensitive res
562	5	2.3	131	1	R14350	HCV protease seque	635	5	2.3	139	1	Y04886	Mycobacterium spec
563	5	2.3	132	1	R68540	Hepatitis C virus	636	5	2.3	140	1	Y04891	Mycobacterium spec
564	5	2.3	133	1	R89746	AFT-1 interacting	637	5	2.3	141	1	R30454	C242:11 Mab kappa
565	5	2.3	134	1	R01690	HCV protease clone	638	5	2.3	141	1	R32541	C242 kappa chain v
566	5	2.3	135	1	W46390	Amino acid sequenc	639	5	2.3	141	1	W00105	Generic ob protein
567	5	2.3	136	1	W97602	Amino acid sequenc	640	5	2.3	141	1	W00050	N-terminally trunc
568	5	2.3	137	1	W00617	Large monomer of m	641	5	2.3	141	1	R91882	Keratinocyte growt
569	5	2.3	138	1	W14582	Streptococcus pneu	642	5	2.3	141	1	R94407	Met-KGF delta-N23/
570	5	2.3	139	1	W53996	Anti-CD4 antibody	643	5	2.3	141	1	R94404	Met-KGF delta-N23/
571	5	2.3	140	1	Y12924	Amino acid sequenc	644	5	2.3	141	1	R94405	Met-KGF delta-N23/
572	5	2.3	141	1	P60017	Sequence of cardio	645	5	2.3	141	1	R94406	Met-KGF delta-N23/
573	5	2.3	142	1	W50977	Norwalk virus RNA-	646	5	2.3	141	1	W00353	KGF delta-N23/R(14
574	5	2.3	143	1	R00108	Generic N-terminal	647	5	2.3	141	1	W61424	Keratinocyte growt
575	5	2.3	144	1	W00052	N-terminally trunc	648	5	2.3	142	1	P61688	Sequence of enzyme
576	5	2.3	145	1	W60949	Streptococcus pneu	649	5	2.3	143	1	P93370	Protein sequence e
577	5	2.3	146	1	Y12656	Human 5' EST seque	650	5	2.3	143	1	P50063	Soybean heat shock
578	5	2.3	147	1	R12298	Norwalk virus RNA-	651	5	2.3	143	1	W98594	H. pylori GHPO 115
579	5	2.3	148	1	W00618	Large monomer of m	652	5	2.3	144	1	R06431	N-terminal of SP1-
580	5	2.3	149	1	W80674	S. pneumoniae prot	653	5	2.3	144	1	R47247	Fragment of 101 kd
581	5	2.3	150	1	Y12005	Human 5' EST seque	654	5	2.3	144	1	W00049	N-terminally trunc
582	5	2.3	151	1	Y04862	Mycobacterium spec	655	5	2.3	144	1	W03523	Generic anti-obesi
583	5	2.3	152	1	P51238	Sequence of pro-at	656	5	2.3	144	1	W03524	Anti-obesity prote
584	5	2.3	153	1	W00069	C-terminally trunc	657	5	2.3	144	1	W47082	Salmonella Sef14 p
585	5	2.3	154	1	W76346	Papillomavirus mai	658	5	2.3	144	1	W98379	H. pylori GHPO 143
586	5	2.3	155	1	W22848	Mortierella alpina	659	5	2.3	145	1	W00541	Human mature obesi
587	5	2.3	156	1	W79301	A Staphylococcus a	660	5	2.3	145	1	W00302	Human delta Gln28
588	5	2.3	157	1	W71300	Obesity protein an	661	5	2.3	145	1	W03002	Human cystatin F.
589	5	2.3	158	1	W00107	Generic N-terminal	662	5	2.3	145	1	W32323	Mature human cysta
590	5	2.3	159	1	W00051	N-terminally trunc	663	5	2.3	145	1	W30893	Synthetic obesity
591	5	2.3	160	1	W14583	Streptococcus pneu	664	5	2.3	146	1	W00539	Human mature obesi
592	5	2.3	161	1	W86277	Blastx output of h	665	5	2.3	146	1	W00013	Acid stable modifi
593	5	2.3	162	1	R62446	BLC6 light chain v	666	5	2.3	146	1	R99490	Chimeric ob protei
594	5	2.3	163	1	Y10859	Amino acid sequenc	667	5	2.3	146	1	R99491	Chimeric ob protei
595	5	2.3	164	1	R33951	CTMO1 VL Anti-hum	668	5	2.3	146	1	R99492	Chimeric ob protei
596	5	2.3	165	1	W00619	Large monomer of m	669	5	2.3	146	1	R99493	Chimeric ob protei
597	5	2.3	166	1	W27117	Feline interleukin	670	5	2.3	146	1	R99494	Chimeric ob protei
598	5	2.3	167	1	W29751	Anti-HMEG MAB CTMO	671	5	2.3	146	1	R99495	Chimeric ob protei
599	5	2.3	168	1	Y11226	S. pneumoniae prot	672	5	2.3	146	1	W00014	Chimeric ob protei
600	5	2.3	169	1	P90154	Sequence of hepati	673	5	2.3	146	1	W00015	Chimeric ob protei
601	5	2.3	170	1	P92037	Sequence encoded i	674	5	2.3	146	1	W00016	Chimeric ob protei
602	5	2.3	171	1	P70564	Product of ORF 6 f	675	5	2.3	146	1	R99489	Generic ob protein
603	5	2.3	172	1	R09420	B38.1 Light Chain	676	5	2.3	146	1	R99499	Acid stable modifi
604	5	2.3	173	1	W06209	MAB B38-1 light ch	677	5	2.3	146	1	W00010	Acid stable modifi
605	5	2.3	174	1	W14585	Streptococcus pneu	678	5	2.3	146	1	W00011	Acid stable modifi
606	5	2.3	175	1	W85056	Mouse B38.1 light	679	5	2.3	146	1	R99500	Acid stable modifi
607	5	2.3	176	1	W00106	Generic N-terminal	680	5	2.3	146	1	W00012	Acid stable modifi
608	5	2.3	177	1	W00061	N-terminally trunc	681	5	2.3	146	1	W00301	Human ob protein.
609	5	2.3	178	1	W00620	Large monomer of m	682	5	2.3	146	1	W05524	Wild type ob prote
610	5	2.3	179	1	W15762	Cotton fibrous tis	683	5	2.3	146	1	W22900	Biologically activ
611	5	2.3	180	1	W53672	H. pylori ORF hp0p	684	5	2.3	146	1	W10151	Properly folded ob
612	5	2.3	181	1	P81138	Sequence of plant	685	5	2.3	146	1	W22901	Biologically activ
613	5	2.3	182	1	R54978	Spinach acyl carri	686	5	2.3	146	1	W10152	Properly folded ob
614	5	2.3	183	1	W09638	Oat-derived thioni	687	5	2.3	146	1	W22902	Biologically activ
615	5	2.3	184	1	W28208	Staphylococcus aur	688	5	2.3	146	1	W10153	Properly folded ob
616	5	2.3	185	1	R14542	Hepatitis C Virus	689	5	2.3	146	1	W22903	Biologically activ
617	5	2.3	186	1	R14353	HCV protease seque	690	5	2.3	146	1	W08599	Properly folded ob

691	5	2.3	146	1	W08595	Properly folded ob	764	5	2.3	146	1	W45473	Anti-obesity prote
692	5	2.3	146	1	W22866	Biologically activ	765	5	2.3	146	1	W45474	Anti-obesity prote
693	5	2.3	146	1	W11279	Generic leptin mim	766	5	2.3	146	1	W45476	Anti-obesity prote
694	5	2.3	146	1	W11278	Generic leptin mim	767	5	2.3	146	1	W45478	Anti-obesity prote
695	5	2.3	146	1	W11275	Human leptin for t	768	5	2.3	146	1	W53329	Obesity protein an
696	5	2.3	146	1	W18621	Obesity protein an	769	5	2.3	146	1	W53330	Obesity protein an
697	5	2.3	146	1	W18622	Obesity protein an	770	5	2.3	146	1	W53331	Obesity protein an
698	5	2.3	146	1	W18623	Obesity protein an	771	5	2.3	146	1	W53332	Obesity protein an
699	5	2.3	146	1	W18630	Obesity protein an	772	5	2.3	146	1	W53333	Obesity protein an
700	5	2.3	146	1	W18624	Obesity protein an	773	5	2.3	146	1	W53334	Obesity protein an
701	5	2.3	146	1	W18625	Obesity protein an	774	5	2.3	146	1	W53335	Obesity protein an
702	5	2.3	146	1	W18626	Obesity protein an	775	5	2.3	146	1	W53336	Obesity protein an
703	5	2.3	146	1	W18627	Obesity protein an	776	5	2.3	146	1	W53337	Obesity protein an
704	5	2.3	146	1	W18628	Obesity protein an	777	5	2.3	146	1	W53338	Obesity protein an
705	5	2.3	146	1	W18629	Obesity protein an	778	5	2.3	146	1	W53339	Obesity protein an
706	5	2.3	146	1	W24031	Human obesity prot	779	5	2.3	146	1	W53340	Obesity protein an
707	5	2.3	146	1	W13838	Chimpanzee leptin	780	5	2.3	146	1	W53341	Obesity protein an
708	5	2.3	146	1	W13839	Gorilla leptin. Re	781	5	2.3	146	1	W53342	Obesity protein an
709	5	2.3	146	1	W13840	Orangutan leptin.	782	5	2.3	146	1	W50999	Anti-obesity prote
710	5	2.3	146	1	W30727	Haemoglobin beta c	783	5	2.3	146	1	W51000	Anti-obesity prote
711	5	2.3	146	1	W32568	Anti obesity prote	784	5	2.3	146	1	W51001	Anti-obesity prote
712	5	2.3	146	1	W32569	Anti obesity prote	785	5	2.3	146	1	W51002	Anti-obesity prote
713	5	2.3	146	1	W32570	Anti obesity prote	786	5	2.3	146	1	W51003	Anti-obesity prote
714	5	2.3	146	1	W32571	Anti obesity prote	787	5	2.3	146	1	W51004	Anti-obesity prote
715	5	2.3	146	1	W32572	Anti obesity prote	788	5	2.3	146	1	W51005	Anti-obesity prote
716	5	2.3	146	1	W32573	Anti obesity prote	789	5	2.3	146	1	W50993	Anti-obesity prote
717	5	2.3	146	1	W32574	Anti obesity prote	790	5	2.3	146	1	W50994	Anti-obesity prote
718	5	2.3	146	1	W32575	Anti obesity prote	791	5	2.3	146	1	W50995	Anti-obesity prote
719	5	2.3	146	1	W32576	Anti obesity prote	792	5	2.3	146	1	W50996	Anti-obesity prote
720	5	2.3	146	1	W32577	Anti obesity prote	793	5	2.3	146	1	W50998	Anti-obesity prote
721	5	2.3	146	1	W32578	Anti obesity prote	794	5	2.3	146	1	W57258	Amino acid sequenc
722	5	2.3	146	1	W32579	Anti obesity prote	795	5	2.3	146	1	W62303	Anti-obesity prote
723	5	2.3	146	1	W32580	Anti obesity prote	796	5	2.3	146	1	W62304	Anti-obesity prote
724	5	2.3	146	1	W26189	Obesity protein an	797	5	2.3	146	1	W62305	Anti-obesity prote
725	5	2.3	146	1	W26190	Obesity protein an	798	5	2.3	146	1	W62306	Anti-obesity prote
726	5	2.3	146	1	W26191	Obesity protein an	799	5	2.3	146	1	W62307	Anti-obesity prote
727	5	2.3	146	1	W26192	Obesity protein an	800	5	2.3	146	1	W62308	Anti-obesity prote
728	5	2.3	146	1	W26193	Obesity protein an	801	5	2.3	146	1	W62309	Anti-obesity prote
729	5	2.3	146	1	W26194	Obesity protein an	802	5	2.3	146	1	W62310	Anti-obesity prote
730	5	2.3	146	1	W26195	Obesity protein an	803	5	2.3	146	1	W62601	Formula I for an o
731	5	2.3	146	1	W26196	Obesity protein an	804	5	2.3	146	1	W62602	Formula II for an
732	5	2.3	146	1	W26197	Obesity protein an	805	5	2.3	146	1	W62603	Formula III for an
733	5	2.3	146	1	W26198	Obesity protein an	806	5	2.3	146	1	W62605	Preferred obesity
734	5	2.3	146	1	W26199	Obesity protein an	807	5	2.3	146	1	W62606	Preferred obesity
735	5	2.3	146	1	W26200	Obesity protein an	808	5	2.3	146	1	W62607	Preferred obesity
736	5	2.3	146	1	W26201	Obesity protein an	809	5	2.3	146	1	W62608	Preferred obesity
737	5	2.3	146	1	W34483	Human obesity prot	810	5	2.3	146	1	W62609	Preferred obesity
738	5	2.3	146	1	W34489	Obesity protein an	811	5	2.3	146	1	W62610	Preferred obesity
739	5	2.3	146	1	W34490	Obesity protein an	812	5	2.3	146	1	W62604	Formula IV for an
740	5	2.3	146	1	W34491	Obesity protein an	813	5	2.3	146	1	W49019	Mutant human obese
741	5	2.3	146	1	W34492	Obesity protein an	814	5	2.3	146	1	W49020	Mutant human obese
742	5	2.3	146	1	W34493	Obesity protein an	815	5	2.3	146	1	W49021	Mutant human obese
743	5	2.3	146	1	W34494	Obesity protein an	816	5	2.3	146	1	W49022	Mutant human obese
744	5	2.3	146	1	W34495	Obesity protein an	817	5	2.3	146	1	W62611	Preferred obesity
745	5	2.3	146	1	W34496	Obesity protein an	818	5	2.3	146	1	W57816	Anti-obesity prote
746	5	2.3	146	1	W34482	Human obesity prot	819	5	2.3	146	1	W57817	Anti-obesity prote
747	5	2.3	146	1	W34484	Obesity protein an	820	5	2.3	146	1	W57818	Anti-obesity prote
748	5	2.3	146	1	W34485	Obesity protein an	821	5	2.3	146	1	W57819	Anti-obesity prote
749	5	2.3	146	1	W34486	Obesity protein an	822	5	2.3	146	1	W57813	Anti-obesity prote
750	5	2.3	146	1	W34487	Obesity protein an	823	5	2.3	146	1	W57814	Anti-obesity prote
751	5	2.3	146	1	W34488	Obesity protein an	824	5	2.3	146	1	W57815	Anti-obesity prote
752	5	2.3	146	1	W30791	Obesity protein co	825	5	2.3	146	1	W59919	Leptin receptor ag
753	5	2.3	146	1	W34396	Human Met-OB prote	826	5	2.3	146	1	W70369	Leptin receptor li
754	5	2.3	146	1	W34397	Human Met-OB prote	827	5	2.3	146	1	W69682	Human obesity prot
755	5	2.3	146	1	W30898	Synthetic obesity	828	5	2.3	146	1	W69683	Obesity protein an
756	5	2.3	146	1	W30899	Synthetic obesity	829	5	2.3	146	1	W69684	Obesity protein an
757	5	2.3	146	1	W30900	Synthetic obesity	830	5	2.3	146	1	W69686	Obesity protein an
758	5	2.3	146	1	W36448	Synthetic obesity	831	5	2.3	146	1	W69687	Obesity protein an
759	5	2.3	146	1	W30896	Synthetic obesity	832	5	2.3	146	1	W69688	Obesity protein an
760	5	2.3	146	1	W30897	Synthetic obesity	833	5	2.3	146	1	W69685	Obesity protein an
761	5	2.3	146	1	W30892	Synthetic obesity	834	5	2.3	146	1	W69689	Obesity protein an
762	5	2.3	146	1	W30894	Synthetic obesity	835	5	2.3	146	1	W69690	Obesity protein an
763	5	2.3	146	1	W30895	Synthetic obesity	836	5	2.3	146	1	W69691	Obesity protein an

983 5 2.3 178 1 R14890 Human macrophage c
 984 5 2.3 178 1 R26181 TSG-14. Tumour nec
 985 5 2.3 178 1 R73337 Human interleukin-
 986 5 2.3 178 1 W02202 Human interleukin-
 987 5 2.3 178 1 W60283 Modified xylanase
 988 5 2.3 178 1 W60742 Xylanase I of Tric
 989 5 2.3 178 1 W69734 Human cystatin-lik
 990 5 2.3 179 1 R26529 Fetal calf TP-like
 991 5 2.3 180 1 R63284 Polypeptide encode
 992 5 2.3 180 1 R63277 Polypeptide encode
 993 5 2.3 180 1 R63278 Polypeptide encode
 994 5 2.3 180 1 R63279 Polypeptide encode
 995 5 2.3 180 1 R63280 Polypeptide encode
 996 5 2.3 180 1 R63281 Polypeptide encode
 997 5 2.3 180 1 R63282 Polypeptide encode
 998 5 2.3 180 1 R63283 Polypeptide encode
 999 5 2.3 180 1 W55490 H. pylori ORF hp5p
 1000 5 2.3 180 1 W47081 Salmonella Sef14 p

ALIGNMENTS

RESULT 1
 ID W55509 standard; Protein; 192 AA.
 AC W55509;
 DT 30-JUN-1998 (first entry)
 DE H. pylori ORF 04ge10816_22086531_f2_10 cytoplasmic protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW bacterium.
 OS Helicobacter pylori.
 PN W09737044-A1.
 PD 09-OCT-1997.
 PF 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI; 97-503122/46.
 DR N-PSDB; V24918.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claim 14; Page 584-585; 1145pp; English.
 CC This sequence is a H. pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The DNA and probes derived from it may be used for the
 CC identification of H. pylori in a sample, and the diagnosis of
 CC H. pylori infection. Nucleic acid sequences complementary to the
 CC DNA act as antisense sequences, and can be used to prevent the
 CC translation of H. pylori mRNA. Antibodies against the protein can
 CC be used in immunoassays to evaluate the abundance and distribution
 CC of H. pylori-specific antigens. The genomic sequence of H. pylori
 CC (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were
 CC analysed for ORF of at least 180 nucleotides, and the predicted
 CC coding regions defined by computer evaluation. To identify likely
 CC H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having
 CC identified and determined the sequences of interest, particular
 CC regions can be isolated from H. pylori by PCR amplification for
 CC recombinant polypeptide production, e.g. in E. coli hosts.
 SQ Sequence 192 AA;

Query Match 3.2%; Score 7; DB 1; Length 192;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RGVTLN 134
 DB 4 RGVTLN 10
 |||||

RESULT 2
 ID W55364 standard; Protein; 208 AA.
 AC W55364;
 DT 17-JUN-1998 (first entry)
 DE H. pylori ORF 04ge11210orf1 protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW Helicobacter pylori.
 PN W09737044-A1.
 PD 09-OCT-1997.
 PF 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI; 97-503122/46.
 DR N-PSDB; V24773.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claim 14; Page 584-585; 1145pp; English.
 CC This sequence is a H. pylori protein of unspecified function.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SQ Sequence 208 AA;

Query Match 3.2%; Score 7; DB 1; Length 208;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RGVTLN 134
 DB 20 RGVTLN 26
 |||||

RESULT 3
 ID W87996 standard; Protein; 241 AA.
 AC W87996;
 DT 15-APR-1999 (first entry)

DE A human MCG18 protein.
 KW MCG4 protein; gene regulatory function; heat shock protein;
 KW guanine nucleotide exchange factor protein; MCG7 protein;
 KW heat shock-binding protein; MCG18 protein; zinc finger protein;
 KW cancer.
 OS Homo sapiens.
 PN W09853061-A1.
 PD 26-NOV-1998.
 PF 22-MAY-1998; AU0380.
 PR 22-JAN-1998; AU-001460.
 PR 23-JAN-1997; AU-006972.
 PR 23-MAY-1997; AU-006973.
 PR 23-MAY-1997; AU-006974.
 PR 22-JAN-1998; AU-001458.
 PR 22-JAN-1998; AU-001459.
 PA (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
 PI Gartside M, Grimmond S, Hancock J, Hayward N, Silins G;
 DR WPI: 99-070146/06.
 DR N-PSDB; X04554.
 PT New gene-expression regulatory genes, mcg4, mcg7, and mcg18 - encode
 PT a zinc finger protein, a GEF, and a heat shock or heat shock binding
 PT protein, useful to detect and treat cancer
 PS Claim 7; Fig 19; 80pp; English.
 CC The present sequence represents a MCG18 protein. The protein has gene
 CC regulatory functions, and has homology to a zinc finger protein.
 CC The specification also describes MCG4, which is homologous to
 CC guanine nucleotide exchange factor protein, and MCG7, which is
 CC homologous to a heat shock protein or heat shock-binding protein.
 CC Detection of mutations in the MCG genes can be used to identify the
 CC propensity for various types of cancer, and to treat, arrest, or
 CC otherwise ameliorate, the effects of a cancer in an animal or bird.
 SQ Sequence 241 AA;

Query Match 3.2%; Score 7; DB 1; Length 241;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LWRNPP 158

||||||
 DB 12 LWRNPP 18

RESULT 4
 W11482 ID W11482 standard; Protein; 339 AA.
 AC W11482;
 DT 09-NOV-1998 (first entry)
 DE Helicobacter polypeptide GHPO 408.
 KW GHPO 408; infection; therapy; diagnosis; vaccine; gastritis;
 KW ulcer.
 OS Helicobacter pylori.
 PN W09821225-A1.
 PD 22-MAY-1998.
 PF 14-NOV-1997; U21353.
 PR 29-JUL-1997; US-7902615.
 PR 14-NOV-1996; US-749051.
 PR 01-APR-1997; US-831309.
 PR 01-APR-1997; US-833457.
 PR 01-APR-1997; US-834705.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (LINR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C,
 PI Odenbreit S, Tomb J;
 DR WPI: 98-297855/26.
 DR N-PSDB; V52017.
 PT Helicobacter polynucleotide and polypeptide sequences - useful to
 PT treat or prevent gastrointestinal infection
 PS Claim 1; Page 100-101; 362pp; English.
 CC This claimed Helicobacter pylori polypeptide, designated GHPO 408,
 CC can be used in vaccination methods for preventing or treating

CC Helicobacter infection. 85 Helicobacter polypeptides (see
 CC W71474-W71558) are claimed, as well as isolated polynucleotides
 CC (see V52009-93) that encode them. The invention also provides:
 CC methods for producing these Helicobacter polypeptides in
 CC recombinant host systems, and related expression cassettes, vectors
 CC and transfected or transduced host cells; live vaccine vectors
 CC that contain the polynucleotides of the invention and which can be
 CC used to prevent or treat Helicobacter infection; therapeutic and/or
 CC prophylactic methods involving administration of polynucleotide
 CC molecules, polypeptides or monospecific antibodies; methods for
 CC detecting the presence of Helicobacter in samples using e.g.
 CC the polypeptides or monospecific antibodies; and methods for
 CC purifying the polypeptides by antibody-based affinity
 CC chromatography.
 SQ Sequence 339 AA;

Query Match 3.2%; Score 7; DB 1; Length 339;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RNVGTLN 134

||||||
 DB 151 RNVGTLN 157

RESULT 5

W22307 ID W22307 standard; Protein; 352 AA.
 AC W22307;
 DT 18-MAR-1998 (first entry)
 DE Lipase modulator from P. wisconsinensis.
 KW Pseudomonas wisconsinensis; lipase modulator; expression system;
 KW enzyme production; detergent.
 OS Pseudomonas wisconsinensis.
 PN W09713847-A1.
 PD 17-APR-1997; BE0109.
 PF 14-OCT-1996; BE0109.
 PR 12-OCT-1995; BE-000851.
 PA (GEMV) GENECOR INT INC.
 PI Andre C, Charmolille L, Cornelis P, Dhaese P, Hazbon MH;
 DR WPI: 97-235883/21.
 DR N-PSDB; T73345.
 PT Expression system containing regulatory sequences from P.
 PT wisconsinensis - and optionally modulator and GPW protein sequences,
 PT especially for lipase production for use e.g. in detergents
 PS Claim 23; Pages 36-38; 74pp; French.
 CC This sequence, isolated from Pseudomonas wisconsinensis, is a lipase
 CC modulator. A claimed expression system for production of an enzyme,
 CC especially a Pseudomonas lipase, comprises at least sequences for a
 CC promoter, signal sequence, mature enzyme and a terminator. The
 CC transformed cells are used to produce enzymes, especially lipase for use
 CC in detergents. Other enzymes are useful in food, pharmaceutical and
 CC chemical industries. The system provides effective secretion of large
 CC quantities of enzyme into the culture medium. GPW seems to assist
 CC expression of enzyme and to protect it (and cells) against oxygen
 CC radicals, also to repair cell membranes damaged by oxygen radicals.
 SQ Sequence 352 AA;

Query Match 3.2%; Score 7; DB 1; Length 352;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 YTLDSLA 210

||||||
 DB 203 YTLDSLA 209

RESULT 6

W55695 ID W55695 standard; Protein; 398 AA.
 AC W55695;

DT 07-JUL-1998 (first entry)
 DE H. pylori ORF 13ae10610.859692.c2.32 cytoplasmic protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN WO9737044-A1.
 PD 09-OCT-1997.
 PF 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR N-PSDB; V25104.
 DT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14,96; Pages 957-958; 1145pp; English.
 CC This sequence is a H. pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds. The
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SQ Sequence 398 AA;

Query Match 3.2%; Score 7; DB 1; Length 398;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RNVGTLN 134
 DB 210 RNVGTLN 216
 |||||

RESULT 7

R44434
 ID R44434 standard; Protein; 416 AA.

DT 06-JUN-1994 (first entry)
 DE Subtilisin-like serine protease PEPP from *Aspergillus niger*.
 KW PEPP; serine protease; subtilisin; filamentous fungus;
 KW protease-deficient yeast strain.
 OS *Aspergillus niger* (strain M400).
 PN AU9336959-A.
 PD 21-OCT-1993.
 PF 15-APR-1993; 036959.
 PR 15-APR-1992; EP-810281.
 PR 12-MAR-1993; GB-005097.
 PA (CIBA) CIBA GEIGY AG.
 PI Buxton F;
 DR WPI; 93-386961/49.
 DR N-PSDB; Q52536.

PT New mutant subtilisin-type serine protease(s) derived from

PT *Aspergillus niger* - used to transform hosts to confer reduced
 PT protease activity, for improved yields in protein expression
 PS Claim 17; Page 70-72; 85pp; English.
 CC The pepD gene was isolated from *A.niger* N400 and sequenced. The gene
 CC codes for a subtilisin-type protease PEPP. Engineered *A.niger*
 CC mutants which are unable to produce functional PEPP can be used for
 CC increased recombinant production of heterologous proteins.
 SQ Sequence 416 AA;

Query Match 3.2%; Score 7; DB 1; Length 416;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LPAGIER 145
 DB 75 LPAGIER 81
 |||||

RESULT 8

R84345
 ID R84345 standard; Protein; 764 AA.

AC R84345;
 DT 25-MAR-1996 (first entry)
 DE Melon aconitase encoded by cDNA clone 16.
 KW Arabidopsis thaliana; aconitase; exon; intron; probe; melon; Zea mays;
 KW Cucumis melo; maize; plant metabolism; Krebs cycle; glyoxylate cycle;
 KW citrate; acetyl CoA; catabolism; polysaccharide; lipase; chimeric;
 KW resistance marker; hormone; enzyme; ss.
 OS Cucumis melo.
 PN WO9520046-A1.
 PD 27-JUL-1995.
 PF 25-JAN-1995; E00263.
 PR 25-JAN-1994; FR-000787.
 PA (BIOC-) BIOCEM SA.
 PI Alric M, Perez P, Peyret P;
 DR WPI: 95-269459/35.
 DR N-PSDB; 102387.
 DT New plant aconitase, its fragments and related nucleic acid - also
 PT chimeric genes, transgenic plants, antibodies etc., used to modify
 PT plant metabolism by regulating carboxylic acid prodn.
 PS Example 4; Fig 4; 122pp; French.

CC The amino acid sequence of the melon aconitase encoded by the melon
 CC aconitase gene clone 16. The gene was isolated by screening a melon cDNA
 CC library with antibodies raised against the purified aconitase from melon
 CC seeds. The melon seed aconitase was compared to the potato (*S.tuberosum*)
 CC mitochondrial aconitase and good homology in the N-terminal sequence was
 CC observed. A 2.2 kb EcoRI fragment of the melon aconitase cDNA was used as
 CC a probe to isolate the sequence of the Arabidopsis thaliana aconitase
 CC cDNA sequence (T02366). The A.thaliana gene sequence was then used to
 CC obtain a partial gene sequence of the maize aconitase gene (T02365). The
 CC aconitase genes can be used to modify plant metabolism by overexpression
 CC of aconitase. This leads to overproduction of acids in the Krebs and
 CC glyoxylate cycles, esp. citrate. Fragments of the genes can be used to
 CC inhibit the expression of aconitase, resulting in overproduction of
 CC acetyl CoA, and alterations of metabolism/catabolism of polysaccharides,
 CC lipases, and N cpds. Chimeric genes contg. the aconitase gene or its
 CC promoter, are used to provide controlled expression e.g. during
 CC development, or to express heterologous enzymes, desired traits,
 CC resistance markers, hormones, etc., in plants.
 SQ Sequence 764 AA;

Query Match 3.2%; Score 7; DB 1; Length 764;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 KLRNGVT 132
 DB 120 KLRNGVT 126
 |||||

RESULT 9

OM of: US-09-252-691-7056 to: GenEmbl.* out_format : pfs
 Date: Jun 10, 2000 12:46 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+2n.model -DEV=xlpl
 -Q=/cgn2_1/USPTO_spool/US09252691/runat_05062000_101650_595/app_query.fasta.1
 -DB=GenEmbl -QFWT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINWATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=4.500
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFWT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000 -USER=US09252691
 -NCPY=6 -ICPY=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-252-691-7056
 Query length: 222
 Database: GenEmbl.*
 Database sequences: 882769
 Database length: 486395729
 Search time (sec): 563.060000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_ba2:AE0000213	-	983.00	1434.19	7.9e-72	10959
gb_ba1:AE0000213	-	983.00	1434.19	1.1e-71	15007
gb_ba1:D90748	-	983.00	1428.92	1.5e-71	20284
gb_ba1:D90749	-	850.00	1212.68	1.7e-59	180993
gb_ba1:AC022161	-	824.00	1173.81	2.5e-57	186591
gb_ba1:AC009127	-	695.50	982.80	1.1e-46	228805
gb_ba1:AC008911	-	598.00	861.81	6.0e-40	11833
gb_ba2:U32752	-	474.50	686.49	3.5e-30	5476
gb_ba2:SLF250721	-	474.50	686.49	1.4e-28	133859
gb_ba1:D90899	-	298.00	396.10	5.2e-14	123469
gb_ba1:SYCSLRA	-	272.00	379.14	4.6e-13	9999
gb_ba2:AE0001942	-	264.50	362.27	4.0e-12	19331
gb_ba2:AE0001708	-	259.00	356.32	8.6e-12	15077
gb_ba1:AE0000741	-	253.00	342.05	5.4e-11	28206
gb_ba1:BA00013	-	253.00	342.05	5.4e-11	28206
gb_ba1:BSUB00113	-	245.00	348.24	2.4e-11	3415
gb_ba2:ECU18111	-	245.00	348.24	2.4e-11	3415
gb_ba2:AE0000225	-	245.00	337.59	9.5e-11	11850
gb_ba1:D90764	-	245.00	334.72	1.4e-10	16373
gb_ba1:D90765	-	245.00	333.57	1.6e-10	18962
gb_ba2:U32799	-	244.50	337.41	9.7e-11	11093
gb_ba2:U32799	-	236.50	302.10	9.0e-09	171311
gb_ba1:MTG1125	-	233.50	310.67	3.0e-09	37432
gb_ba1:AC021060	-	224.00	280.11	1.5e-07	255800
gb_ba1:AC021060	-	218.50	287.67	5.7e-08	40745
gb_ba1:MLCB1351	-	217.50	286.57	6.6e-08	38936
gb_ba1:U00021	-	214.00	281.32	1.3e-07	39193
gb_ba1:RPXK03	-	209.50	257.83	2.6e-06	379110
gb_ba1:AR007490	-	206.00	291.26	3.6e-08	3061
gb_ba1:AE000308	-	206.00	279.59	1.6e-07	11961
gb_ba1:AE000308	-	206.00	269.45	5.9e-07	39149
gb_ba1:BSUB0016	-	203.50	254.10	4.2e-06	123640
gb_ba2:AF008220	-	203.50	253.92	4.3e-06	220060
gb_ba2:AF008220	-	201.50	273.86	3.4e-07	10703
gb_ba2:AE000475	-	201.50	249.89	7.3e-06	176195
gb_ba2:ECU0089	-	201.50	249.89	7.3e-06	176195
gb_ba2:AC010542	-	201.50	247.51	9.9e-05	232632
gb_ba2:AC010542	-	201.50	245.58	1.3e-05	232632
gb_ba2:AF0052507	-	193.00	269.16	6.1e-07	15385
gb_ba1:ECICRCHES	-	187.00	266.60	8.5e-07	3061
gb_ba2:U32804	-	186.00	249.53	7.6e-06	12481
gb_ba2:AE0001667	-	176.50	234.19	5.4e-05	20386
gb_ba2:AE0001667	-	176.50	234.19	5.4e-05	20386
gb_ba2:AE0018634	-	176.50	207.78	0.0016	315277

gb_ba1:BATROPEA - 175.50 237.34 3.6e-05 8392 ! Z19055 B.aphidicola tryptop
 gb_ba1:MX99740 - 166.00 221.38 0.0003 10414 ! X99740 Mycoplasma hominis
 gb_ba1:RP282292 - 163.00 246.39 1.1e-05 333 ! Z82292 R.prowazekii genomic

seq_name: gb_ba2:AE0000213

seq_documentation_block: 10959 bp DNA BCT 12-NOV-1998
 LOCUS AE0000213 Escherichia coli K-12 MG1655 section 103 of 400 of the complete genome.

ACCESSION AE000213 U00096

VERSION AE000213.1 GI:1787371

KEYWORDS Escherichia coli.

SOURCE Escherichia coli

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 10959)
 Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,Kiley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.

The complete genome sequence of Escherichia coli K-12

JOURNAL Science 277 (5331), 1453-1474 (1997)

MEDLINE 97426617

REFERENCE 2 (bases 1 to 10959)

AUTHORS Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 3 (bases 1 to 10959)

AUTHORS Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 4 (bases 1 to 10959)

AUTHORS Plunkett, G. III.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K-12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K-12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

Location/Qualifiers

source

1. 10959


```

/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
13..40
/promoter
protein_bind
/feature="factor Sigma70; predicted +1 start at 1184959"
119..136
/feature="central position to predicted promoter:82"
135..1381
/feature="PhoB predicted site"
155..1381
/feature="pept"
/feature="bl127"
155..1381
/feature="pept"
/EC_number="3.4.11.-"
/function="putative enzyme; Not classified"
/feature="0408; 100 pct identical to 42 aa fragment
PEPT_ECOLISW: P29745; 93 pct identical to 407 residues of
409 aa PEPT_SALTY SW: P26311"
/codon_start=1
/transl_table=11
/product="putative peptidase T"
/protein_id="AAC74211.1"
/db_xref="GI:1787372"
/db_xref="PID:g1787372"
/translation="MDKLLERFLNVSLDTOSKAGVROVPSTEGQKLLHLKLEQLEE
MGLINVTISEKGLMATLPANVPDIPAIIGTISHVDTSFDCSGKNNPQIVENTYRGDD
TALGIGDVSFVPEVPLVQLLQTLITDGLKADKAGIAEINTALAVLOQKI
PHGDIRFATDEEVGKAKHFDVDAFWAYTVGGGVGELEFENFAASVNIKIV
GNVHPFAGKGVNALSIAARIEAEVPADESPEMTEGEGYFHLASMKGVRAADWH
YIIRDFRQFQKRRKMKMETAKIKVGLHPCDYLIVIEDSYNNREKVVVEHPHIL
IAQAMRDCDIEPELPIRGCTDGAQLSFMGLPCPNLFTGGYNYHGKHEFTLEGMEK
AVOVIVTAELTAQRK"
complement(1430..2560)
/gene="ycfB"
/gene="bl128"
complement(1430..2560)
/feature="ycfB"
/function="orf; Unknown"
/feature="f376; 100 pct identical to fragment YCFD_ECOLI
SW:P27431 but has 3 additional N-terminal aa and 93
additional C-terminal aa"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74212.1"
/db_xref="GI:1787373"
/db_xref="PID:g1787373"
/translation="MLNNEYQLTNWPDFLERHWKRPVVLKRGFNFIPIGPDELA
GLAMESVDSRLVSHQCKQVSHGPFESYDHLGETNWSLLVOAVNHWHPETALMRP
FRELDPNRIDDLMTSFSVPGGVPHLDQYDVFTIOCTGRRRVFCVKLQMKQCHRP
DLQVDPEAIIDELSPGDILATIPRGPHGVALENAMNYSVGFRAFRNRELISGFA
DYVLQRELGNYSDDVPVRAHPADVLPOEMKREMLLELQNPHEFKQWFGFIS
QSRHELIDIAPEPPYQDEIYDALKQGEVLRVGLGRLVRIGDDVYANGEKIDSPHRP
ALDALASNLTAENFNGDALEDFSLAMLAALVNSGYWFFEG"
complement(2593..2613)
/feature="factor Sigma70; predicted +1 start at 1187488"
complement(2627..4087)
/gene="phoQ"
/gene="bl129"
complement(2627..4087)
/feature="phoQ"
/EC_number="2.7.3.-"
/function="enzyme; Global regulatory functions"
/feature="f486; 99 pct identical to PHOQ_ECOLI SW: P23837"
/codon_start=1
/transl_table=11
/product="sensor protein PhoQ"
/protein_id="AAC74213.1"
/db_xref="GI:1787374"
/db_xref="PID:g1787374"
/translation="MKLLRLFPFLSLRVFLATAAVVLVLSLAYGMVALIGYSVSF
DKTFFLLRGESNLFYTLAKWENKHLVPELNIDKQSPMTLTLYDENGQLLWQARDY

```

```

PMLMKIQPDLKSNKGFEIEADVNDTSLLSGDHSIQOOLQOEVEREDDDAEMTHSYA
VNVYPATSRMPKLTIVVDITPVELKSSYMWMSWFIYVLSANLLVILPLAAVWNSL
RPIELAKEVRELLEHNRLLNPATRELISLVNRLNRLKSERDYKTYRTTUDLT
HSLTPIAVLQSTLRSKMSVSDAEVPMLEQISRSIQOIGYILHRAVMRGSTLLS
RELHPAPLNDLNTSALKVYQKRGVNSIDISPEISFQEQNDVFVGMVGLDNACK
YCLFEFVSARQTDHELYIVVEDDGPGLPSKREVIIDRGORVDTLRPGQGVGLAVAR
EITQYEGKIVAGESMLGARMGVIFGRQHSAPKDE"
complement(4087..4758)
/gene="phoP"
/feature="phoP"
/feature="bl130"
complement(4087..4758)
/feature="phoP"
/function="regulator; Global regulatory functions"
/feature="f223; 100 pct identical to PHOP_ECOLI SW: P23836"
/codon_start=1
/transl_table=11
/product="transcriptional regulatory protein"
/protein_id="AAC74214.1"
/db_xref="GI:1787375"
/db_xref="PID:g1787375"
/translation="MRVLVVEDNALLRHLLKVOIQDAGHOVDDAEADKADYVINEHI
PDIAIVDGLPDEDEGLSLIRWSNDVSLPILVLTARESQDKVEVLSAGADVTRK
PHIEVYMARQALMRRNSGLASQVISLPFQVDSRLSINDEVIKLTAFETYIMET
LIRNKGVSQKSLMLQLYPDALRESHTIDVLMRLKRIQAIQYQEVITTVRGQY
LFLER"
complement(4257..4287)
/gene="phoP"
/feature="phoP"
/feature="factor Sigma70; predicted +1 start at 1189162"
complement(4927..6297)
/gene="purB"
/feature="purB"
complement(4927..6297)
/feature="purB"
/EC_number="4.3.2.2"
/function="enzyme; Purine ribonucleotide biosynthesis"
/feature="f456; 99 pct identical to PUR8_ECOLI SW: P25739"
/codon_start=1
/transl_table=11
/product="adenylosuccinate lyase"
/protein_id="AAC74215.1"
/db_xref="GI:1787376"
/db_xref="PID:g1787376"
/translation="MELSLTAVSPVDGRYDKVKSALRGIFSEYGLLKRFVQVEVRL
QKLAHAALKEVPFAFAADYGLDAIVASFEEDAAIKTIERTNHDVKAVEYFLKE
KYAIEPELHVSSEFIFHACTSEINNLSHMLKTDARDEVILPYRQLIDGKDLAVQ
YRDIPLLSRTHGPAFTPIGKEMANVAVMERQVRLNQVLEILCKINGAVGNNAHI
AAYPEVDWHQFSEFVTSIGIQWNPYTOIEPHDYIAELFDCVAPRNTLIDFDRDWW
GYIALNHFKTKTAGEIGSTWPHKVNPIDEFSEGNLSNAVLOHLASKLPVSRKQ
RLDTSTVLNRNGIGYALIAIYQSTLKGVSLEVNRLHLLDELHNNWEVLAEPQTV
MRRYGLEKPEYKELKELTRGKRVDAEGMKQFIDGLALPEEEKARLKAMTPANTYIGRAIT
MYDELK"
complement(6301..6942)
/feature="ycfC"
/feature="bl132"
complement(6301..6942)
/feature="ycfC"
/function="orf; Unknown"
/feature="f213; 99 pct identical to YCFC_ECOLI SW: P25746"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"

```

```

alignment_scores:
  Quality: 983.00      Length: 218
  Ratio: 4.795      Gaps: 0
  Percent Similarity: 94.037      Percent Identity: 85.780

alignment_block:
  US-09-252-691-7056 x AE000213/rev
  Align seg 1/1 to reverse of: AE000213 from: 1 to: 10959
  1 AlaileMetArgGlnLeuIleThrProGluAsnThrMethThrylsrHse 17

```


|||||
9268 GCATATATGCGGCAATTCATATCTCGAAATACCATGCAAAATCTTC 9219
17 rPheArgLysHisArgValGluArgPheSerArgGlnAlaThrArgA 34
|||||
9218 TTTTAGAATACACAGGTTAAGCAATTCAGCTCGCAACGTTCTACCAAGC 9169
34 rgThrProGluProGlnProThrArgValIleLeuPheAsnLysProTyr 50
|||||
9168 GTAACCTGAAACACAGCCACCGCTGTGATCTGTTCAATAACCCCTAC 9119
51 AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAs 67
9118 GATGTTCTCCGAGTTCACGATGAGCCGCGACGCAAAACATAAAGA 9069
67 pPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAsps 84
|||||
9068 ATTATCCCGGTTTCAGGTTGTTATGACGACGAGTGCCTTGACCGGATA 9019
84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
|||||
9018 GCGAAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8969
101 ThrGluProGlyLysArgThrGlyLysIleTyrValGlnValGluG1 117
8968 ACCAGCGCGGTAAACGACCGCAAAATCTATATGTCGAGGTGGAAG 8919
117 YGluProAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuA 134
|||||
8918 TATCCACACAGACGACCTTGAAGCCTTGGCAATGGCGTAACCTTAA 8869
134 snAspGlyProThrLeuProAlaGlyIleGluArgValAsnGluProGlu 150
8868 ATGATGCGCCCTACCTCGCGCGCGCGAGCTGTGACGAACCGCG 8819
151 TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProTh 167
8818 TGTTATGCGCGCGGAATCCACCAATTCGTGAACGCAAAAGTATCCAC 8769
167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
8768 CAGCTGCTGAAGTACCTTATATGAGACGCTAATCGCAGGTGCGCC 8719
184 rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla 200
8718 GCATGACCGCCATGTGGCTTCCCGACGCTGCGACTGATCGTATGCG 8669
201 MetGlySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgAspVa 217
8668 ATGGGTGATTACTCTTTGGATAATCTTGGCAATGGTGAATGGCGAGAAGT 8619
217 lThr 218
|||||
8618 GACA 8615
g_name: gb_bai:D90748

seq documentation block:
LOCUS D90748 15007 bp DNA BCT 07-FEB-1999
DEFINITION Escherichia coli genomic DNA. (25.6 - 25.9 min).
ACCESSION D90748 AB001340
VERSION D90748.1 GI:1651553
KEYWORDS Complete and shotgun sequencing; potB; potA; pepT; phoP;
purB; ycfC; ycfB; icdA; icd; icdE; lit.
SOURCE Escherichia coli (strain K12) DNA, clone: Kohara clone #239.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 15007)
AUTHORS Mori, H.
DIRECT SUBMISSION
TITLE Submitted (29-JUL-1996) to the DDBJ/EMBL/GenBank databases.
JOURNAL Hirotada Mori, NARA Institute of Science and Technology, Res. &

Edu. Center for Genetic Info.: 8916-5 Takayama, Ikoma, Nara 630-01,
Japan (E-mail: hmori@gtc.aist-nara.ac.jp, tel: 81-7437-2-5660,
fax: 81-7437-2-5669)
2 (sites)
Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T.,
Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T.,
Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.
The systematic sequencing of the Escherichia coli genome in Japan
Unpublished (1996)
3 (sites)
Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nashimoto, H.,
Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K.,
Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T.
A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)
97061202
Collaboration Information:
Project:
The Japan E. coli genome DNA sequencing project
Group:
The Japan E. coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N.,
Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology,
Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http://bsw3.aist-nara.ac.jp.
Location/Qualifiers
1. 15007
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"
/clone="Kohara clone #239"
/map="25.6-25.9 min"
/note="Nucleotide position 1185468-1200474 from the
initiation site of ThrA (0 min.); This clone is from
Kohara lambda miniset library."
complement(1. 678)
/gene="potB"
complement(<1. 678)
/gene="potB"
/note="ORF ID: o238#14; similar to SwissProt Accession
Number P23860"
/transl_table=1
/product="Spermidine/putrescine transport system permease
protein PotB."
protein_id="BAA35947.1"
/db_xref="GI:1651554"
/translation="MIVTIVGLVFLVFLPNLMIIGTSFLTRDDASFKVKKVFTLDNYT

RLDPLDYFVLLHSLNMLATATLACVLGYPPFAWELAKLPHKVRPLLLFLLIYVPFTN
SLRIYGLKFLSTKGLNFEFLNLWGLVIDPDIRIMFTPPSAVIIGLVYLLPMMVPIY
SSIEKLDKPLEAARDUGAKLQTFIRIIILPLTPIAGCLLMLPAMGLFYVSDLM
GGAKNLI"

gene

complement(692.1828)

CDS

complement(692.1828)

/gene="pota"
/gene="pota"
/note="ORF_ID:0239#1; similar to PIR Accession Number
A40840"

/codon_start=1

/transl_table=11

/product="Spermidine/putrescine transport protein A"

/protein_id="BAA35948.1"

/db_xref="GI:1651553"

/translation="MGOSKLNKPSSLPVLQAGIRKFDGKEVIPQLDLINNGE
FLTLAPGGCGKTVLRIAGLETVDGSRIMLQNETIHPVAENRYVNTVFQYALFP
HMTVEFNAVGLRMOKTPAAEITPRVMEALRMVQLETFAQRKPHOLSGGQOQVATAR
AVNKPRLIDLSLADYKLRQONELKALQRLGITFFVTHDQREALTMSBRI
VWARDRIQDGPRIEYEPKFLVAGFGEINFNATVIERLDQVRVANYEGREC
NIYVFAVEPGQLHLVRPDLRVEIDNDNHAEGLIGYVRERNYKGMTLESVLE
NGKVMVSEFFNEDDPDFDHSLDQKMAINNVESVEVYLADEEHK"

gene

2078.3304

CDS

/gene="pept"
2078.3304

/gene="pept"
/note="ORF_ID:0239#2; similar to SwissProt Accession
Number P29745"

/codon_start=1

/transl_table=11

/product="Peptidase T (EC 3.4.11.-) (aminotripeptidase)

/protein_id="BAA35949.1"

/db_xref="GI:1651556"

/translation="MOKLRFNYSLDTSQAGVRQVPSTEGQKLLHLKEQLEE
MGLINTVLSKGTLMATLPANVPDIPAIQFIISHVDTSPDCSGKNVNPQIVENVRGD
IALGIDVLSVPMFVLHOLLGOTLITDTGKLLGADKAGIAEIMTALVQQKKI
PHGDIRVATPDEVGKAGKFDVADFAKATVDCGGVGELEFENFNANIKIV
GNVHPGTAGVMVNLAAHAEVPADESPETEGYEGFYHLASMKGTVBRADMH
YIIRDFRQKFEARKKMEIAKVKGLHPDCYIELVIEDSYNNRKYVEHPHIL
IAQAMRDCIEPELAPIRGGTGAQLSEFGLPCPLNFTGGYVHGKHEVTEGMEK
AVQIVRIAEITAQRK"

gene

complement(3353.6010)

CDS

/gene="phoQ"
complement(3353.4474)

/gene="phoQ"
/note="ORF_ID:0239#3; similar to PIR Accession Number
D41966"

/codon_start=1

/transl_table=11

/product="orf 2 downstream of phoQ."

/protein_id="BAA35950.1"

/db_xref="GI:4062693"

/translation="MEVQLTNLPDFLERHWKRPVVVLRKGFNFIDTISPDELGLA
MESEVSRVSHODGKQVSHGSPESYDHGLETNWSLLVQAVNWHPEPAAALMRPRE
LPWRIDDLMSFVPCGGVGHLDQDVFIOTGRRVRVGEKQMKQHCPPHLL
QVDPFEAIDEELEGDILYIPPGFHEGVALENAMNYSVGFAPNTRELISFADYLL
LQRELGNNYSYSDPVPRAHPADVLPQEMDKLREMLLEINQPEHFKQWFEISQSR
HELDIAPPEYQDPDEIYDALKQGVLRGLGLRVLIRIGDVTYANGKEIKDSPRHPALD
ALASNTALAEFGDALEDPSFLAMLAALVNSGYWFFEG"

CDS

complement(4550.6010)

/gene="phoQ"
/note="ORF_ID:0239#4; similar to PIR Accession Number
B41966"

/codon_start=1

/transl_table=11

/product="Virulence membrane protein phoQ."

/protein_id="BAA35951.1"

/db_xref="GI:1651557"

/translation="MKRLLLFFPLSLRVRFLLATAAVLVLSLAYGMVALIGYSVSF
DKTTFRLRGESNLFYTLAKWNNKLHVELPENIDKQSPMTLIYDENGOLLWAQRDV
PWLKMIQPLWKSNGFHEAEADVNDTSLLSGDHSIQOQLQEVREDDDAETHSVA
VNYVPATSRPKLTIVVDIPIVELKSSYMWWSFYVLSANLLLIPLLWVAWWSL

RPIEALAKEVRELEHNRELLNPATRTRELTSLVRNRLRLKSEREDKYRTLTDLT
HSUKTPLAVLQSTLRSEKMSVSDAEPVMEQISRIQQIYIYLRASMRGTLLS
RELPYAPLLDNLTSALNKYQKGNISLIDISPEISFGEQNDVFYVGMVGLDNACK
YCLFEYVEISARQTDHLYIVVEDDGPGLSKREVIFDRGQVRVDTLRPGQGVGLAVAR
EITEQYEGKIVAGESMLGARMEVIFGRQHSAPKDE"

gene

complement(6010.6681)

CDS

complement(6010.6681)

/gene="phop"
/gene="phop"
/note="ORF_ID:0240#1; similar to SwissProt Accession
Number P23836"

/codon_start=1

/transl_table=11

/product="Transcriptional regulatory protein Phop."

/protein_id="BAA35952.1"

/db_xref="GI:1651558"

/translation="MNVLVEDNALLRHLKVIQIDAGHQVDDAEADKADYVYLNH
PDIAIVDLGLPDEDGLSLRRWSNDVSPILVLTARESWQDKVEVLSAGADDYVTKP
FHEEYMARQALMRNSGLASQVLSLPPFQVDSLRSRELSINDEVIKLTAFETIMET
LIRNNKGVSKDSLMQLPYDAELRESHTIDVLMGLRKKIKQAQYPOEVITTVRGQGY
LFELR"

gene

complement(6850.8220)

CDS

complement(6850.8220)

/gene="purB"
/note="ORF_ID:0240#2; similar to PIR Accession Number
S19212"

/codon_start=1

/transl_table=11

/product="Adenylosuccinate lyase (EC 4.3.2.2)"

/protein_id="BAA35953.1"

/db_xref="GI:1651559"

/translation="MELSLTAVSPVDGRYDKVSALRGIFSEYGLLKFRVQVEVRWL
QKLAHAALKESEFPAFAADAIGYDAIVASFEEDAARIKTERTNHDVKAVYFLKE
KVAIEPELHAVSEFHFACSTDENNLSHMLKTARDEVILPYWQLIDGKDLAVQ
YRDIPLLSTHQCPATPSIGEMANVAMEROYQRLQVQVEILKINGAVGYNNAHI
NAYPEWQFSEEFYSLGIONPITQIDEPHDYIAELFCVARENTILIDFDROVW
GYALNHFQKTKIAGEISGTMPHKYNPIDFENSEGNLGLSNVLAHLASKLPVSRWQ
RDLTDSVTLNRLGVGIGYALIAQSTLKGVSLEVRNHLDELHWNVEVLAEPVQTV

alignment_scores:

Quality: 983.00 Length: 218

Ratio: 4.795 Gaps: 0

Percent Similarity: 94.037 Percent Identity: 85.780

alignment_block:

US-09-252-691-7056 x D90748/rev ..

Align seg 1/1 to reverse of: D90748 from: 1 to: 15007

1 AlarileMetArgGinLeuileThrProGluAsnThrMetThrLysThrSe 17
|||||GCTATAATGCGCAATTCATAATCTCTGAAAATACCATGCAAAAACCTC 11142
17 rPheArgLysHisArgValGluArgPheSerSerArgGlnAlaThrArgA 34
|||||TTTTAGAAATCACCAGGTAGCGGATTCAGTCGCAAGCTTCTACGAGC 11092
34 rGThrProGluProGlnProThrArgValIleLeuPheAsnLysProThr 50
|||||GTAAACCTGAAACACCGCCAGCGGTGATCCTGTTCAATAAACCCTAC 11042
51 AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAs 67
|||||GATGTTCTCCGAGTTCACCGATGAAGCCGACGCAAAACATTAAAGA 10992
67 pPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAsp 84
10991 ATTCATCCCGGTTTCAGGTGTTTATGACGAGGTGCGCTTGACCGCGATA 10942
84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
10941 CGGAAGGTTTCTGCTGCTGACCAATACCGCGGCTTGACGCGCGGTITA 10892


```

S19212"
/codon_start=1
/transl_table=11
/product="Adenylosuccinate lyase (EC 4.3.2.2)"
/protein_id="BAA35962.1"
/db_xref="GI:1651565"
/translation="MELSLTAVSPVDGRGVKVSALRGIFSEYGLLKRQVQVEVRL
KLAHAARKEVFAFAADAGYDAIVASFEEDAAIRKTIERTNHDVKAVEYFLKE
QVAEIPHLAVSFIFHACFSEDINNLSHALMLKTAARDEVILPYWQLIDGKDLAVQ
YRIPLLSRTHGQAPATSTIGKMANVAYRMEQYRQLNOVEILGKINGAVGYNNAHI
ANPEVDWHQFSEFFYSLGIONPVTQIPEHDYIAELFDCVARENTILIDFDRDVM
GYTALNFKORTAGEISSTMEHKVNPIDFENSEGNLGSNAVLOHLSKLPVSRWQ
RDTDSTVLNKGVIQYALIAVOSTLKGVSKEIVNRDHLDELNDHNEVLEPIQTV
MRYGDIKPYEKLKELTRGRVDAEGMKQPIDGLALPEEKARLKAMTPANYIGRAIT
MVDELK"
complement(2541..3182)
/gene="ycfc"
/gene_cofc="
complement(2541..3182)
/feature="ORF_ID:0240#3; similar to PIR Accession Number
S19211"
/codon_start=1
/transl_table=11
/product="Hypothetical protein 23."
/protein_id="BAA35963.1"
/db_xref="GI:4062714"
/translation="MAKNYDITLALAGICQARLVQOLAHQHCADALHVSLSNII
DMNPSTLAVFGSEANLRVGLLETLGLVNASRQGLNAELTRYTLISLMYLERKLSSA
KGLADTNGRNGRQROLEHFDLOSETLSMAAAIYVDVISPGLPQVGTGSPAVLQS
POVQKVRATLLAGIRAVLWHQVGGRLQLMFSSRLTLTQAKQILAHUTPEL"
complement(3218..4324)
/gene="ycfb"
complement(3218..4324)
/gene_cofb="
/feature="ORF_ID:0240#4; similar to SwissProt Accession
Number P25745"
/codon_start=1
/transl_table=11
/product="Hypothetical protein in purB 5' region (orf-15)
."
/protein_id="BAA35964.1"
/db_xref="GI:4062715"
/translation="MSTAKKVIYMGSGVDSSVALLQOQGVQVEGLEFMKWEEDD
GEETCAADLADQAQVCDKLGIELHYVNFANEDVNDVFLFAEYKAGRTNPDILC
NKEIKFALEFAEDLGADYIAEHVYRADVDGKRLLRLDSDKQDSFYLYTLSH
EQIAQSLFPVGELEKPKQVRIAEGLVLTAKKDSGICFIGERKREFGLRILPAQP
KIIITVQDEIGEHOGLMYHTLQORGLGIGGTEGTEEPWYVYVDKDVNNILVVAQG
HEHPLMSVGLIAQOLHWVDREPTGTMRCTVTKRYRQTDIPCTVKALDDDRIEVIFD
EPVAATPGQSAFYNGEVCILGGGIIQRPLPV"
complement(4378..4839)
/feature="ORF_ID:0240#5; similar to PIR Accession Number
S64074"
/codon_start=1
/transl_table=11
/product="Hypothetical protein YGL067w"
/protein_id="BAA35965.1"
/db_xref="GI:4062716"
/translation="MFKPHVTVACVHAEGKFLVVEETINGKALWNOPAGHLEADEFL
VEAARELWEETGISAPQFIRHWOIAPDKTFLRFLFAIELEQICPTQPHDSID
CCRWVSAEIIQAQNLSPVAESIRCYQSQRYPLEMIGDFNWPFTKGI"
complement(4849..5502)
/feature="ORF_ID:0240#6; similar to PIR Accession Number
I64156"
/codon_start=1
/transl_table=11
/product="Hypothetical protein HI0694"
/protein_id="BAA35966.1"
/db_xref="GI:4062717"
/translation="MRQFIISNTWQTSFRNHQKFRPSQSRTRRKPENQPRVILF
NKPYDVLFOFDEKRLTKREFIPVQGYAAGRLDRDSEGLLVTNNGALQARLTQPG
KRTKIYVQVEGPTODALELRNGVTLNDGPTLPAGAEVLVDEPAWLPNRPRIER
KSIPTSLKILTYLGRNRQVRMTAHVGFPLRLIRYAMGDYSLDNLANGEWREVTD"
5674..6924

```

```

/feature="icd"
5674..6924
/feature="icd"
/feature="ORF_ID:0240#7; similar to SwissProt Accession
Number P08200"
/codon_start=1
/transl_table=11
/product="Isocitrate dehydrogenase (NADP) (EC 1.1.1.42)
(oxalosuccinate decarboxylase) (idh) (NADP+-specific icdh)
(idp)."
/protein_id="BAA35967.1"
/db_xref="GI:1651566"
/translation="MESKVVYPAQKKIILQNGKLVNENPIPIESDGIQVDYTPA
MLKVVDAAVEKAYRGERKISWMEIYTGKSTQVQGVQVWLPAETLDLIREYVAIKGP
LTPPYGGIRISNVALRQELDIYICLRPVRYQGPSPVKHPKELDMVIFRENSDIY
AGIEWKADSADAERKVKIFREMGVKIRFPEHCIGIKPCSEEGTKRLVRAAIEYAI
ANDRDSVTLVHKGNIMKFTGAFKDWQYQALAREEFGGELIDGGPWLKVNPNPTGKEIV"
alignment_scores:
Quality: 983.00 Length: 218
Ratio: 4.795 Gaps: 0
Percent Similarity: 94.037 Percent Identity: 85.780
alignment_block:
US-09-252-691-7056 x D90749/rev
Align seg 1/1 to reverse of: D90749 from: 1 to: 20284
1 AlaileMetArgGlnLeuIleThrProGluAsnThrMetThrLysThrse 17
5508 GCTATATGCGGCAATTCATAATCTCTGAAATACCATGCAAAAACCTC 5459
17 rPheArgLysHisArgValGluArgPheSerSerArgGlnAlaThrArga 34
5458 TTTTGAATAATCACCAGGTTAAGCATTAGCTCAGTCGCAACGTTCTACCA 5409
34 rGThrProGluProGlnProThrArgValIleLeuPheAsnLysProTyr 50
5408 GTAAACCTGAAACACGACCCACGCGTGTGATCCTGTTCAATAAACCCCTAC 5359
51 AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAs 67
5358 GATGTTCTTCGCGAGTTCACCGCATGAAGCGGACGCAAAACATTAAAA 5309
67 pPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAsps 84
5308 ATTATCCCGGTTTCAGGGTGTATTATGCAGCAGGTGCGCTTGACCGCGATA 5259
84 rGluGlyLeuLeuValLeuThrAsnAspGlyValValLeuGlnAlaArgLeu 100
5258 CGGAAGGTTTCTGGTGTGACCAATACGCGCGGTTCAGCGCGCTTTA 5209
101 ThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGl 117
5208 ACCCAGCGGGTAAACGACCGCAAAATCTATTATGTGCAAGTGAAGG 5159
117 yGluProAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeu 134
5158 TATTCCACACAAAGCAGCACTTGAAGCCTTCGCAATGGCGTAAACCTTAA 5109
134 snAspGlyProThrLeuProAlaGlyIleGluArgValAsnGluProGlu 150
5108 ATGATGGCCCTTACCTTGGCCCGCGCGGAGCTGGTTGACGAACCCGCG 5059
151 TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProTh 167
5058 TGGTATTGGCGCGGAATCCACCAATTCGTGAACGCAAAAGTATTCCCA 5009
167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArg 184
5008 CAGCTGGCTGAAGTACCTTATATGAAGGACGTAATGCCAGGTGGGCC 4959
184 rGMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla 200

```

|||||
4958 GCATGACCGCCATGTTGGCTTCCCGACGTCGACGATTCGCTATGCG 4909

201 MetGlySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgAspVa 217

|||||
4908 ATGGGTGATTAATCTTTGGATAAATCTGCCAATGGTGAATGCCGAGAGT 4859

217 lThr 218

|||||

4858 GACA 4855

seq_name: gb_htg7:AC022161

seq_documentation_block:

LOCUS AC022161 180993 bp DNA HTG 26-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-273P11, LOW-PASS SEQUENCE
SAMPLING.

ACCESSION AC022161

VERSION AC022161.1 GI:6758614

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 180993)

DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 16

REFERENCE 2 (bases 1 to 180993)

DOE Joint Genome Institute.

Direct Submission

Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

* NOTE: This record contains 113 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 511: contig of 511 bp in length

gap of unknown length

512 1507: contig of 996 bp in length

gap of unknown length

1508 2279: contig of 772 bp in length

gap of unknown length

2280 2614: gap of unknown length

gap of 335 bp in length

2615 3450: contig of 836 bp in length

gap of unknown length

3451 3914: contig of 464 bp in length

gap of unknown length

3915 4858: contig of 944 bp in length

gap of unknown length

4859 5380: contig of 522 bp in length

gap of unknown length

5381 6247: contig of 867 bp in length

gap of unknown length

6248 7203: contig of 956 bp in length

gap of unknown length

7204 7920: contig of 717 bp in length

gap of unknown length

7921 8924: contig of 1004 bp in length

gap of unknown length

8925 9674: contig of 750 bp in length

gap of unknown length

9675 11049: contig of 1375 bp in length

gap of unknown length

11050 11834: contig of 785 bp in length

gap of unknown length

11835 12774: contig of 940 bp in length

gap of unknown length

12775 13555: contig of 781 bp in length

gap of unknown length

13556 14538: contig of 983 bp in length

gap of unknown length

14539 15398: contig of 860 bp in length

gap of unknown length

15399 15696: contig of 298 bp in length

gap of unknown length

15697 16675: contig of 979 bp in length

gap of unknown length

16676 17597: contig of 922 bp in length

gap of unknown length

17598 18969: contig of 1372 bp in length

gap of unknown length

18970 19300: contig of 331 bp in length

gap of unknown length

19301 19380: contig of 80 bp in length

gap of unknown length

19381 20457: contig of 1077 bp in length

gap of unknown length

20458 21187: contig of 730 bp in length

gap of unknown length

21188 21435: contig of 248 bp in length

gap of unknown length

21436 22621: contig of 1186 bp in length

gap of unknown length

22622 23209: contig of 588 bp in length

gap of unknown length

23210 23994: contig of 785 bp in length

gap of unknown length

23995 24643: contig of 649 bp in length

gap of unknown length

24644 25421: contig of 778 bp in length

gap of unknown length

25422 25826: contig of 405 bp in length

gap of unknown length

25827 27532: contig of 1706 bp in length

gap of unknown length

27533 28937: contig of 1405 bp in length

gap of unknown length

28938 30011: contig of 1074 bp in length

gap of unknown length

30012 30574: contig of 563 bp in length

gap of unknown length

30575 31545: contig of 971 bp in length

gap of unknown length

31546 32314: contig of 769 bp in length

gap of unknown length

32315 33010: contig of 696 bp in length

gap of unknown length

33011 33654: contig of 644 bp in length

gap of unknown length

33655 34705: contig of 1051 bp in length

gap of unknown length

34706 35755: contig of 1050 bp in length

gap of unknown length

35756 36672: contig of 917 bp in length

gap of unknown length

36673 37904: contig of 1232 bp in length

gap of unknown length

37905 38457: contig of 553 bp in length

gap of unknown length

38458 39301: contig of 844 bp in length

gap of unknown length

39302 40257: contig of 956 bp in length

```
*      40258      gap of unknown length
*      41691:      contig of 1434 bp in length
*      43227:      contig of unknown length
*      43228      contig of 1536 bp in length
*      44416:      gap of unknown length
*      44417      contig of 1189 bp in length
*      45206:      gap of unknown length
*      45207      contig of 790 bp in length
*      46223:      gap of unknown length
*      46224      contig of 1017 bp in length
*      47104:      gap of unknown length
*      47105      contig of 881 bp in length
*      49150:      gap of unknown length
*      49151      contig of 2046 bp in length
*      50227:      gap of unknown length
*      50228      contig of 1077 bp in length
*      51990:      gap of unknown length
*      51991      contig of 1763 bp in length
*      52651:      gap of unknown length
*      52652      contig of 661 bp in length
*      54157:      gap of unknown length
*      54158      contig of 1506 bp in length
*      55684:      gap of unknown length
*      55685      contig of 1527 bp in length
*      56936:      gap of unknown length
*      56937      contig of 1252 bp in length
*      57993:      gap of unknown length
*      57994      contig of 1057 bp in length
*      58775:      gap of unknown length
*      58776      contig of 882 bp in length
*      60147:      gap of unknown length
*      60148      contig of 1272 bp in length
*      61314:      gap of unknown length
*      61315      contig of 1167 bp in length
*      61502:      gap of unknown length
*      61503      contig of 188 bp in length
*      62298:      gap of unknown length
*      62299      contig of 796 bp in length
*      63529:      gap of unknown length
*      63530      contig of 1231 bp in length
*      63537      gap of unknown length
*      63577      contig of 1847 bp in length
*      65461:      gap of unknown length
*      65462      contig of 85 bp in length
*      66798:      gap of unknown length
*      66799      contig of 1337 bp in length
*      68065:      gap of unknown length
*      68066      contig of 1267 bp in length
*      69742:      gap of unknown length
*      69743      contig of 1677 bp in length
*      71264:      gap of unknown length
*      71265      contig of 1522 bp in length
*      72180:      gap of unknown length
*      72181      contig of 916 bp in length
*      74061:      gap of unknown length
*      74062      contig of 1881 bp in length
*      75047:      gap of unknown length
*      75048      contig of 986 bp in length
*      76521:      gap of unknown length
*      76522      contig of 1474 bp in length
*      78628:      gap of unknown length
*      78629      contig of 2107 bp in length
*      80347:      gap of unknown length
*      80348      contig of 1719 bp in length
*      82199:      gap of unknown length
*      82200      contig of 1852 bp in length
*      83804:      gap of unknown length
*      83805      contig of 1605 bp in length
```

```
alignment_scores:
  Quality: 850.00      Length: 193
  Ratio: 4.749        Gaps: 1
  Percent Similarity: 92.746      Percent Identity: 85.492
```

alignment_block:

US-09-252-691-7056 x AC022161

Align seg 1/1 to: AC022161 from: 1 to: 180993

```
1 AlaleMetArgGlnLeuLleThrProGluAsnThrMetThrLysThrSe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31737 GCTATAATGCGCAATCATATAATCTCTGAAATACCATGCAAAACTTC 31786

17 rPheArgLysHisArgValGluArgPheSerSerArgGlnAlaThrArgA 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31787 TTTTAGAAATCACCAGGTATAGCGATTACGCTCGCAACGTTCTACAGGC 31836

34 rGThrProGluProGlnProThrArgValLleLeuPheAsnLysProTyr 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31837 GTCAACCTGAAACACGAGCCAGCGGTGTGATCCTGTTCATAAACCTTAC 31886

51 AspValLeuProGlnProThrArgPheSerSerArgGlnAlaThrLysAs 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31887 CATGTTCTTCGCGAGTTTCAACGATGAAGCCGCGACGCAAAACATTAAAGA 31936

67 pPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAsps 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31937 ATTCAATCCGGTTTCAAGGTTTATGACGAGGTGCGCTTGACCGCGATA 31986

84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31987 CGGAAGGGTCTGCTGTGTGACCAATAACGCGCGGTTCAGGCGCGTTTA 32036

101 ThrGlnProGlyLysArgThrGlyLysIleTyrValGlnValGluG1 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32037 ACCGACGCGGTAAACGACCGGAAATCTATTATGTGACAGTGGGAGG 32086

117 y-GluProAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeu 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32087 TATTCCTCCACACAGACGACTTGAAGCCTTGCCTGCAATGGCGTAACCTTA 32136

134 AsnAspGlyProThrLeuProAlaGlyLleGluArgValAsnGluProG1 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32137 AATGATGCGCTTACCTGCGCGCGCGGAGCTGTTGACGAAACCCGC 32186

150 uTrpLeuTrpProArgAsnProIleArgGluArgLysSerIleProt 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32187 GTGGTTATGCGCGGGAATCCACCAATTCGTGACGCAAAAGTATCCCA 32236

167 hrSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArg 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32237 CCAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCCGAGGTGCGC 32286

184 ArgMetThrAlaHisValGlyPhePro 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32287 CGCATGACCGCCCATGTTGGCTTCCCC 32313
```

seq_name: gb_htg7:AC009127

seq_documentation_block:

```
LOCUS      AC009127      186591 bp      DNA      HTG      26-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-498D10, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION      AC009127
VERSION      AC009127.2      GI:6758895
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 186591)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 186591)
AUTHORS      DOE Joint Genome Institute.
```

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 26, 2000 this sequence version replaced gi:5685948.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

* NOTE: This record contains 101 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 989: contig of 989 bp in length
* gap of unknown length
* 990 1318: contig of 329 bp in length
* gap of unknown length
* 1319 1843: contig of 525 bp in length
* gap of unknown length
* 1844 2066: contig of 223 bp in length
* gap of unknown length
* 2067 3127: contig of 1061 bp in length
* gap of unknown length
* 3128 3535: contig of 408 bp in length
* gap of unknown length
* 3536 4171: contig of 636 bp in length
* gap of unknown length
* 4172 4797: contig of 626 bp in length
* gap of unknown length
* 4798 5170: contig of 373 bp in length
* gap of unknown length
* 5171 5390: contig of 220 bp in length
* gap of unknown length
* 5391 6021: contig of 631 bp in length
* gap of unknown length
* 6022 6686: contig of 665 bp in length
* gap of unknown length
* 6687 6975: contig of 289 bp in length
* gap of unknown length
* 6976 7331: contig of 356 bp in length
* gap of unknown length
* 7332 7911: contig of 580 bp in length
* gap of unknown length
* 7912 8248: contig of 337 bp in length
* gap of unknown length
* 8249 8844: contig of 596 bp in length
* gap of unknown length
* 8845 9149: contig of 305 bp in length
* gap of unknown length
* 9150 9469: contig of 320 bp in length
* gap of unknown length
* 9470 9911: contig of 442 bp in length
* gap of unknown length
* 9912 10494: contig of 583 bp in length
* gap of unknown length
* 10495 11143: contig of 649 bp in length
* gap of unknown length
* 11144 11762: contig of 619 bp in length
* gap of unknown length
* 11763 12380: contig of 618 bp in length
* gap of unknown length
* 12381 13000: contig of 620 bp in length
* gap of unknown length
* 13001 13576: contig of 576 bp in length
* gap of unknown length
* 13577 14175: contig of 599 bp in length
* gap of unknown length
*
* 14176 14801: contig of 626 bp in length
* gap of unknown length
* 14802 15373: contig of 572 bp in length
* gap of unknown length
* 15374 15716: contig of 343 bp in length
* gap of unknown length
* 15717 16489: contig of 773 bp in length
* gap of unknown length
* 16490 17111: contig of 622 bp in length
* gap of unknown length
* 17112 17855: contig of 744 bp in length
* gap of unknown length
* 17856 19178: contig of 1323 bp in length
* gap of unknown length
* 19179 20052: contig of 874 bp in length
* gap of unknown length
* 20053 20790: contig of 738 bp in length
* gap of unknown length
* 20791 21464: contig of 674 bp in length
* gap of unknown length
* 21465 22165: contig of 701 bp in length
* gap of unknown length
* 22166 22966: contig of 801 bp in length
* gap of unknown length
* 22967 23458: contig of 492 bp in length
* gap of unknown length
* 23459 24138: contig of 680 bp in length
* gap of unknown length
* 24139 24984: contig of 846 bp in length
* gap of unknown length
* 24985 25642: contig of 658 bp in length
* gap of unknown length
* 25643 26565: contig of 923 bp in length
* gap of unknown length
* 26566 27196: contig of 631 bp in length
* gap of unknown length
* 27197 27815: contig of 619 bp in length
* gap of unknown length
* 27816 27901: contig of 86 bp in length
* gap of unknown length
* 27902 28512: contig of 611 bp in length
* gap of unknown length
* 28513 29881: contig of 1369 bp in length
* gap of unknown length
* 29882 30600: contig of 719 bp in length
* gap of unknown length
* 30601 31204: contig of 604 bp in length
* gap of unknown length
* 31205 32333: contig of 1129 bp in length
* gap of unknown length
* 32334 32911: contig of 578 bp in length
* gap of unknown length
* 32912 33816: contig of 905 bp in length
* gap of unknown length
* 33817 34778: contig of 962 bp in length
* gap of unknown length
* 34779 36264: contig of 1486 bp in length
* gap of unknown length
* 36265 38613: contig of 2349 bp in length
* gap of unknown length
* 38614 39144: contig of 531 bp in length
* gap of unknown length
* 39145 41224: contig of 2080 bp in length
* gap of unknown length
* 41225 41982: contig of 758 bp in length
* gap of unknown length
* 41983 43378: contig of 1396 bp in length
* gap of unknown length
* 43379 44856: contig of 1478 bp in length
* gap of unknown length
* 44857 46513: contig of 1657 bp in length
* gap of unknown length

```
* 46514 48278: contig of 1765 bp in length
* gap of unknown length
* 48279 49123: contig of 845 bp in length
* gap of unknown length
* 49124 50185: contig of 1062 bp in length
* gap of unknown length
* 50186 51798: contig of 1613 bp in length
* gap of unknown length
* 51799 52794: contig of 996 bp in length
* gap of unknown length
* 52795 54154: contig of 1360 bp in length
* gap of unknown length
* 54155 56271: contig of 2117 bp in length
* gap of unknown length
* 56272 58993: contig of 2722 bp in length
* gap of unknown length
* 58994 60268: contig of 1275 bp in length
* gap of unknown length
* 60269 62633: contig of 2365 bp in length
* gap of unknown length
* 62634 65099: contig of 2466 bp in length
* gap of unknown length
* 65100 66776: contig of 1677 bp in length
* gap of unknown length
* 66777 68874: contig of 2098 bp in length
* gap of unknown length
* 68875 72266: contig of 3392 bp in length
* gap of unknown length
* 72267 74691: contig of 2425 bp in length
* gap of unknown length
* 74692 77588: contig of 2897 bp in length
* gap of unknown length
* 77589 80107: contig of 2519 bp in length
* gap of unknown length
* 80108 84825: contig of 4718 bp in length
* gap of unknown length
* 84826 87029: contig of 2204 bp in length
* gap of unknown length

alignment_scores:
  Quality: 824.00      Length: 188
  Ratio: 4.763         Gaps: 0
  Percent Similarity: 92.021  Percent Identity: 84.043

alignment_block:
US-09-252-691-7056 x AC009127 ..
Align seq 1/1 to: AC009127 from: 1 to: 186591
1 AlaileMetArgGlnLeulleThrProGluAsnThrMetThrLysThrse 17
27948 GCTATAATGCGCAATTCATAATCTCTGAAATACCATCAAAAAACTTG 27997
17 rPheArgLysHisArgValGluArgPheSerArgGlnAlaThrArgA 34
27998 TTTTAAATACACAGGTTAAGCGATTACAGTCGCACACGTTCTACCGC 28047
34 rGthrProGluProGlnProThrArgValIleLeuPheAsnLysProTyr 50
28048 GTAACCTGAAACACAGCCACCGGTGTGATCTGTCTCAATAACCCCTAC 28097
51 AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLys 67
28098 GATGTTCTCCGAGTTACCGATGAGCGGAGCGGACGAAACATTAAGA 28147
67 pPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAsp 84
28148 ATTATATCCCGTTCAGGGTGTATGACGAGGTGCGCTTGACCGGATA 28197
84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
28198 GCGAAGGGTGTGCTGGTGTGCTACCAACAAACGCGCGTGTGACGGCGGTTA 28247
```

```
101 ThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluG1 117
122248 ACCCAGCGGGTAAACGACCGGAAAAATCTATATGTGAGGTGAAGG 28297
117 yGluProAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuA 134
28298 TATTCCTCCACACAGACGACTTGAAGCCTTGGCAATGGCGTAACCTAA 28347
134 snAspGlyProThrLeuProAlaGlyIleGluArgValAsnGluProGlu 150
28348 ATGATGCGCTTACCTGCGCGCGCGGAGCTGGTTGACGACCGCGG 28397
151 TrpLeuTrpProArgAsnProIleArgGluArgLysSerIleProTh 167
28398 TGGTTATGCGCGCGGAATCCACCAATTCGTGACGCAAAAGTATTCCAC 28447
167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
28448 CAGCTGGCTTAAGATCACCTTATATGAAGACGTAATCGCGAGTTGCGC 28497
184 rgMetThrAlaHis 188
28498 GCATGACGCGCCCAT 28511
seq_name: gb_htg6:AC008911
seq_documentation_block:
LOCUS AC008911.2 226805 bp DNA HTG 14-JAN-2000
DEFINITION Homo sapiens chromosome 5 clone CITB-H1_2268J5, WORKING DRAFT
SEQUENCE, 95 unordered pieces.
ACCESSION AC008911
VERSION AC008911.2 GI:6693307
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 226805)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
2 (bases 1 to 226805)
AUTHORS DOE Joint Genome Institute.
REFERENCE Direct Submission
TITLE DOE Joint Genome Institute.
JOURNAL
COMMENT
-----Genome Center
Center: Joint Genome Institute
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 120029 bases at least Q40
Consensus quality: 163743 bases at least Q30
Consensus quality: 177506 bases at least Q20
Estimated insert size: 226805; sum-of-contigs estimation
Quality coverage: 4.03x in Q20 bases; agarose-fp estimation
Quality coverage: 2.91x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 95 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1018: contig of 1018 bp in length
* gap of unknown length
* 1019 2070: contig of 1052 bp in length
* gap of unknown length
* 2071 3108: contig of 1038 bp in length
* gap of unknown length
```



```
* * *
* 3109 4211: gap of unknown length
* * contig of 1103 bp in length
* 4212 5226: gap of unknown length
* * contig of 1015 bp in length
* 5227 6336: gap of unknown length
* * contig of 1110 bp in length
* 6337 7420: gap of unknown length
* * contig of 1084 bp in length
* 7421 8436: gap of unknown length
* * contig of 1016 bp in length
* 8437 9514: gap of unknown length
* * contig of 1078 bp in length
* 9515 11442: gap of unknown length
* * contig of 1928 bp in length
* 11443 12908: gap of unknown length
* * contig of 1466 bp in length
* 12909 14243: gap of unknown length
* * contig of 1335 bp in length
* 14244 15370: gap of unknown length
* * contig of 1127 bp in length
* 15371 16374: gap of unknown length
* * contig of 1004 bp in length
* 16375 17484: gap of unknown length
* * contig of 1110 bp in length
* 17485 18486: gap of unknown length
* * contig of 1002 bp in length
* 18487 19535: gap of unknown length
* * contig of 1049 bp in length
* 19536 20747: gap of unknown length
* * contig of 1212 bp in length
* 20748 21794: gap of unknown length
* * contig of 1047 bp in length
* 21795 22910: gap of unknown length
* * contig of 1116 bp in length
* 22911 24510: gap of unknown length
* * contig of 1600 bp in length
* 24511 26330: gap of unknown length
* * contig of 1820 bp in length
* 26331 27876: gap of unknown length
* * contig of 1546 bp in length
* 27877 29050: gap of unknown length
* * contig of 1174 bp in length
* 29051 30329: gap of unknown length
* * contig of 1279 bp in length
* 30330 31614: gap of unknown length
* * contig of 1285 bp in length
* 31615 33078: gap of unknown length
* * contig of 1464 bp in length
* 33079 34288: gap of unknown length
* * contig of 1210 bp in length
* 34289 35341: gap of unknown length
* * contig of 1053 bp in length
* 35342 36695: gap of unknown length
* * contig of 1354 bp in length
* 36696 37729: gap of unknown length
* * contig of 1034 bp in length
* 37730 39224: gap of unknown length
* * contig of 1495 bp in length
* 39225 40393: gap of unknown length
* * contig of 1169 bp in length
* 40394 41828: gap of unknown length
* * contig of 1435 bp in length
* 41829 43639: gap of unknown length
* * contig of 1811 bp in length
* 43640 45023: gap of unknown length
* * contig of 1384 bp in length
* 45024 46157: gap of unknown length
* * contig of 1134 bp in length
* 46158 47250: gap of unknown length
* * contig of 1093 bp in length
* 47251 48869: gap of unknown length
* * contig of 1619 bp in length
* * gap of unknown length

* * *
* 48870 50138: contig of 1269 bp in length
* * gap of unknown length
* 50139 51169: contig of 1031 bp in length
* * gap of unknown length
* 51170 52255: contig of 1086 bp in length
* * gap of unknown length
* 52256 53454: contig of 1199 bp in length
* * gap of unknown length
* 53455 55052: contig of 1598 bp in length
* * gap of unknown length
* 55053 56384: contig of 1332 bp in length
* * gap of unknown length
* 56385 57426: gap of unknown length
* * contig of 1042 bp in length
* 57427 58894: contig of 1468 bp in length
* * gap of unknown length
* 58895 59899: contig of 1005 bp in length
* * gap of unknown length
* 59900 61097: contig of 1198 bp in length
* * gap of unknown length
* 61098 62827: contig of 1730 bp in length
* * gap of unknown length
* 62828 64143: contig of 1316 bp in length
* * gap of unknown length
* 64144 65320: contig of 1177 bp in length
* * gap of unknown length
* 65321 66338: contig of 1018 bp in length
* * gap of unknown length
* 66339 67938: contig of 1600 bp in length
* * gap of unknown length
* 67939 69534: contig of 1596 bp in length
* * gap of unknown length
* 69535 70638: contig of 1104 bp in length
* * gap of unknown length
* 70639 71721: contig of 1083 bp in length
* * gap of unknown length
* 71722 72737: contig of 1016 bp in length
* * gap of unknown length
* 72738 74240: contig of 1503 bp in length
* * gap of unknown length
* 74241 76005: contig of 1765 bp in length
* * gap of unknown length
* 76006 77033: contig of 1028 bp in length
* * gap of unknown length
* 77034 79330: contig of 2297 bp in length
* * gap of unknown length
* 79331 80402: contig of 1072 bp in length
* * gap of unknown length
* 80403 82066: contig of 1664 bp in length
* * gap of unknown length
* 82067 83385: contig of 1319 bp in length
* * gap of unknown length
* 83386 85506: contig of 2121 bp in length
* * gap of unknown length
* 85507 87079: contig of 1573 bp in length
* * gap of unknown length
* 87080 88558: contig of 1479 bp in length
* * gap of unknown length
* 88559 89799: contig of 1241 bp in length
* * gap of unknown length
* 89800 92035: contig of 2236 bp in length
* * gap of unknown length
* 92036 94093: contig of 2058 bp in length
* * gap of unknown length
* 94094 96588: contig of 2495 bp in length
* * gap of unknown length
* 96589 98564: contig of 1976 bp in length
* * gap of unknown length
* 98565 100900: contig of 2336 bp in length
* * gap of unknown length
* 100901 103171: contig of 2271 bp in length
* * gap of unknown length
* 103172 105268: contig of 2097 bp in length
```



```
similarity: putative"
/codon_start=1
/transl_table=11
/product="glycerol uptake facilitator protein (glpF)"
/protein_id="AAC22350.1"
/db_xref="GI:1573694"
/translation="MDKSLKANCIGFEFLGTALLIFFGVCVAAALKVAGASGLWEISI
WAGMGVALYATGASGALHNPAVTIALMFACFDGKKVPIYIISQMLGAFFAALY
IALYRNFIDYETHVNIHVRGTOESLSLSTFTSPHPSLGGFAVEFVITAILMAL
IMALTDDGNGRPLAPLLGIIIAVIGGAMGLTFGAMPNADFPGKFFAYLAGWG
ELALTGGREIYFVPMVAPVLGALAGAWLYKKAIGGNLPCNCGE"
960..2471
/gene="HI0691"
CDS
960..2471
/feature="HI0691"
/feature="similar to GB:L19201 SP:P08859 GB:M18393 GB:M55990
GB:X15054 percent identity: 76.89; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="glycerol kinase (glpK)"
/protein_id="AAC22351.1"
/db_xref="GI:1573695"
/translation="WTDKXYIIADOGTSSRAVLIDHNANVVEIAOREFTQIYPRAG
WYEHNPHEIWAQSTLNEYAKAGITSDAIAIGITNQRTTIWKESTGTPYNAI
WQCRTADITDKLKADGHEEYIRNTGLVDDPPIFGTKVKWILDNVEGAREKAE
LLFTVDVFWLWIKTQGVHRYVDYTNASRLFNHITKOWDDKMLIENIPRSMLEP
RNSBIEYQITGKGGVRIYVAGIAGDQQAALXHLCHVAGAKNTYGTGCFMLLHT
GNKAITSKNGLLTIACNAKEPEYAGSVFIAGASIQWLDELKIVHDSFDEYFA
QKVTDNSGVYVPAFTGLGAPYWPYARGAIFGLSRGANRHIVRATLESYATOTRV
LEMQSDSGERLOLYRVDGATNNFLMQFOADILDVNPVRVKEVYALGAAYLAGL
ATGFWQDUDELKARVERFTSPSDNEKERRRKGWKAQRSLEWAKEDDE"
complement(2546..3013)
/gene="HI0692"
complement(2546..3013)
/gene="HI0692"
/feature="similar to SP:P26972 GB:X52093 GB:X63336 PID:47695
percent identity: 74.03; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="xanthine-guanine phosphoribosyltransferase
(gptB)"
/protein_id="AAC22352.1"
/db_xref="GI:1573693"
/translation="MSEKIVVTWDMFQMHARRLSRLLPASQWKGIIVSRGGLFPAA
VLARGLGRHETVCIAHYHDHNNOGELQVLHAAQVPGGEGFIVDDLVDTGNTARA
IRQYYPNAKFVTVPAKPAAGAEVDVVIDIPQNTWIEQPDWDLGLTFVPPLSRK"
3297..4121
/gene="HI0693"
3297..4121
/gene="HI0693"
/feature="similar to GB:M68502 SP:P26093 PID:1573696 percent
identity: 99.64; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="lipoprotein E (hel)"
/protein_id="AAC22353.1"
/db_xref="GI:1573696"
/translation="NKTILKMTALAALSAFVLAGCGSHQKSEGHANNMQLOQAVLGL
NMWQSGEYKALAYQAYNAAKVAFDHAKVAKKAVVADLDETMLDSPAGWQVQN
NKFPDGKQTRWDARQSRVAGFAEFNNYNSHGKVFYVYNNRDKSTEGKTDIDMK
RIGFNGVEESAFYLLKDKSAKAAFAEIEKQGYEIVLVXVGNLDGFGNTVYKGLNADR
RAFVDQNGKFKGTFIMLPNANYGGWEGGLAEGYFKDKTQGIKARLDAVQAWDGK"
4261..4983
/gene="HI0694"
4261..4983
/gene="HI0694"
/feature="similar to GB:U00096 SP:P75966 PID:1787380 percent
identity: 66.29; identified by sequence similarity;
putative"
/codon_start=1
```

```
similarity: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22354.1"
/db_xref="GI:1573697"
/translation="MMLQAFONRNHKLMMKNFSAKGLPSKGSVAFNHFRTFKPLKPK
TSLSDENKVLNFNPFVLTQFTDEQGRATLKDFISIPNVYAGCLDRSDSEGLILITN
NGELQHLRADPKFKTEKYVQVVEGIPETDLAQLRKGVELKGVTKSAKVLISEPN
LWERNPPRERKNIPITWLEIKISEGRNRQVRMTAHIGFPTLRLVRVSMGLLSINGL
ENGFRLLSDEIKALFQTKL"
complement(5109..6080)
/gene="HI0695"
complement(5109..6080)
/gene="HI0695"
/feature="similar to GB:L06129 SP:P29014 PID:147343 GB:U00096
PID:178848 percent identity: 53.62; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="exopolysphatase, putative"
/protein_id="AAC22355.1"
/db_xref="GI:1573698"
/translation="MNDSEILPKHRGNVREIAADLGSNPFHMIVARIVNGSIOVLSR
LKQVKLAEGDENAVLNOEAITRGVNCALFAERLQGFPMENNVVGTVTLRRAVNN
DEFUQAKVPPYPIINISGTEAKTIYAGVCHTQPEKGRKLVLDIGGSGEMIGDD
FIPMAKSRHMGCVSFRAIFDTIDTISFENQRAQSAVNKIEDLGIEIRKLGMQSVL
GSSCTIKRVAQVIATNDPNDPITAEIRNALIEOTLQAKHTEININGLNDQDRDVVF
PGLAILSAVDFVFIQOMRYSDGALREGVIYSLEKNEQVADIRASTA"
complement(6088..9984)
/gene="HI0696"
complement(6088..9984)
/gene="HI0696"
/feature="similar to GB:U14003 SP:P39322 PID:537063 GB:U00096
percent identity: 34.87; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22356.1"
/db_xref="GI:1573699"
/translation="MTEIQIPSETSPKSPKPKHWHVRCIGSAVILVPLGVAG
ALFSDAGOKSLIQLVDMDSFVSEOVGGLQNLVKNRYQTAGIETHIAQARLQL
DSCGLFSREVCRLDFTLNKPTIAINTALLPSPADNSKSGMKRISLISIAENLVM
QDLSVNDQTSITIGNFKSAVSLNNEKGLTIAPTEINDISVIKALSEKSEPKAEQ
NKPVDMAAEQSLTPAFIGNVSEIILPDLHIPEISGNWQYQAVNEKGETLQSVMS
SLIAQADTVNQLOQLAVESLGNSSQGLQDGMPLDLTLKSHLEPLSDGKE
ILPDAVDLTLSSGLKSTALSLTKGVLDALNGVOLAQDKMPLNLTLYAKQYIT
FVNTWPLKINDVTLKLTGLLNVHAELKGDVAGMNYIPASQVFNADGKLYEVTNK
LGISLDGKSEFVGNANWNCANWDIOADLEKMTAFVFPVMPATLSKLSHSGFAGS
QGWQVEVPVADNGLMSAKPISLGSATLNQNLVLTVPDLQIKYGENLKASGVLDH
SDFALDINAPNLRLGSLDKGRVRAISGOITPNDLDTLSSNLHQGFQLAKS
IKGHINNASLSSGLINKAEQHYGGNTKLHLLDLDLSDGQNHKLILKSQGEYVAA
LQINGHFTLEQWKTISQVFFETPIDGVKSQQAIAVSDNKOQTAIAHSCWQNTD
VELCPQAFNAGKOGNIPFQKRYNLDLVNKLIEONSLKNOVGNQVAFWTFDKPQF
TANVDGNHAFSQKLDYETFKLYIPKLTNADIONNNLVLTNDVNHQGRIVGDHIL
NDLAKNRQGLAERLNLSIAQLUTSGESVGEVSVKLSFGNLEKPLNGDFNI
RNIRTKKMPVNITDGDIALFNDRNSTLOGIKRTVDSHLNLRGRANWIEHWTE
LNAQANNFNDIPSMALRFSNPTIRANPELNSGTVDIPWARIKIDSLPDTAEY
SEDEVILNPHKSEIKREFAAKTKSGMEIRDLINIGKDSLDAYGKLTNDGL
LSVKODGNLFGQINLTGRYASFQGLIRKGLISFGQATOPTNIAIRNPET
MEDSKITAGVRVIGIADSEVYIPSEPSKQDQALSYLLTGRSLESSEVGSTGSGA
ALIGLISKSKSLVSGISEGVFIQDLNLGTSGVGDKSKVTYSGNITNRLQIKYGVLF
DGLAEVTURYRLMPOLYFQSVSSNQVDFLLYKFEF"
complement(9994..11730)
```

alignment_scores:
Quality: 598.00 Length: 201
Ratio: 3.891 Gaps: 2
Percent Similarity: 80.597 Percent Identity: 59.204

alignment_block:
US-09-252-691-7056 x U32752

Align seg 1/1 to: U32752 from: 1 to: 11833

COMMENT Potential protein coding regions were assigned on the basis of
similarity search of the ORFs and Genemark analysis.

FEATURES
Location/Qualifiers
1..133859
/organism="Synecocystis sp."
/strain="PC6803"
/db_xref="taxon:1143"
1..772
/gene="ispB"
1..772
/gene="ispB"
/note="ORF_ID:slr0611"
/codon_start=2
/transl_table=11
/product="prenyltransferase"
/protein_id="BAAL6579.1"
/db_xref="GI:1651651"
/translation="AHRRLAEITEMHTASLVHDDVDVDEADLRNVTNLSFDNRV
AVLAGDFLFAQSSWYLANLNEVVKLSEVIRDFAEGEILQISINRDTDDLEYLE
KSYFTASLIANSKAGVLSADPRVCDHLYEYKHLGLAFQIVDDILDFTSTEV
KPGAGSLISGNITAPALFAMEKYPGLKGLIEREFAQAGDLEALVELVEQDGI
ELAAQALRAQHLVLEMSAPRESLLELDVYVLRHL"
937..1494
/note="ORF_ID:slr0612"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAAL6580.1"
/db_xref="GI:1651652"
/translation="MGRDQDSEGLLLTSNKLQHLRAHEFAHQTYFAQVGSPT
DDELEPRGFIADYPTPAIAKLIITPDPPPPNPPYRASIFTSWLSLTLEGRN
RQVRMTAAVGPPTLRVQVQVTRSPQCKGSAATCWLITLGLSPGQWRPLTPW
EENFCQQLTNGPNPQWKKFGDR"
1577..2098
/note="ORF_ID:slr0613"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAAL6581.1"
/db_xref="GI:1651653"
/translation="MSYLIADVAVNRIAAEAYTTLEQAGFAQKNLTICTGYKTADEF
GLVDPKQAKRAKAMWLVPFPAAGYCNLTGLTLDWAGDPGNHIVGLGLGAI
GTMGSFFVGGVGLSFGSDSLPRLNLLQAGKYLVAAGGELQQRATNLLRLPNE
YLGGTAPDEAIV"
complement(2172..2873)
/note="ORF_ID:slr1058"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAAL6582.1"
/db_xref="GI:1651654"
/translation="MSDNLTELSQLDASEKKQLTAAALAEAGGCGILLDYAL
NYPELVAVGNVYQTLNLEQETITQLRNYPGIFPLQSAQIDYLPQEAAGS
QDFEADETDKLCELACPGASQWLYFEVEKFPALDLTINALWHLNSGNFGF
SVQRNLASGEFELKPKICKWKSNGVWTRPKGFTWDLSPAQGHLPILNQLRGV
RSLNHPVWSQGW"
complement(3192..4268)
/gene="pNIL34"
complement(3192..4268)
/gene="pNIL34"
/note="ORF_ID:slr11214"
/codon_start=1
/transl_table=11
/product="pNIL34"
/protein_id="BAAL6583.1"
/db_xref="GI:1651655"
/translation="WNTLEKQGFDEIRPGVTPAKETILTFRYTFDFAEMWDIS
PNEDELRALIERVDYNNHHVRNFSNKHSDHIDGKQRQLPFVEFLERSKAEISGF
LYLEKGRLLKNNKLLAECLNMSDEARHAGFLNKAUSDNLGDLGFLTKSKYT
FFPKPFIFATYLSKIGWRYITTYHLEKNPNDCIYPIFEFFWCDENRHDGFF
DALMRAQPTLNDWAKLWCRFFLLSVFATMINDTQADFYACLGLEARSYDKVIE
KNETAGRVFPIILDVNNPEFYNRLTVCVSNNEQKRAIDASGAPGVKALRKLPIFAS
NGWQFKLYLMKPIADVQLAGAVR"

CDS
complement(4451..5389)
/note="ORF_ID:slr11213"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAAL6584.1"
/db_xref="GI:1651656"
/translation="MLSLENQRILVTGGAGFLGKQVVAOLIAAGDRPKITIPRSKDC
DLQASACERAVENQDVIHLAAHVGIGLNRKPAELFYDNLMMGVQLIHAHQAGV
KKEVCYTICAYPKFTVPFKEDLNWGPETNAPYVAKKALLVQLESYRLOVGFN
GIYLLPNLYGPNEDFDRSSHVIPALIKHVEAQAQKQKQLPVMGDSPTREFLYST
DAARGIVMGTOYKADPYNLTGNTFISIKDLTELICELMEFEEDGIWETDQPNQOPR
RCLDTTKARAEFGFAQVSLRGLKNTIDWYRQRS"
complement(5534..5651)
/gene="rfbD"
complement(5534..5622)
/gene="rfbD"
/note="ORF_ID:slr11212"
/codon_start=1
/transl_table=11
/product="GDP-D-mannose dehydratase"
/protein_id="BAAL6585.1"
/db_xref="GI:1651657"
/translation="MSKSKVLLTGITGDSVLSLELLKGVQVHGIIIRRTSTFNTD
RIDLYVDHDLKALRLHWGLTDCGTLRLRILEDKVPTLYNLQAQSHRVVSDSPE
ITVDSVAMGTLRLLEALRIDQHRIGIQVRYQAGSEMFGKQVEIQKETTPTPRSP
YAKAYGVHQTVYNYRESYDLFACNGILFNHSPRGETFVTRKTRATARIIVAGTOK
KLYGNIDSKRDWGYAKDYVRAMMLQEQDDYVATGETHEVKEFLEIFAFGYVNL
WNQYVAFDRYLRPAEVDLLIGDPKTRKAGWEPSTVITELVHLMVEADLAVGLT
SPNQSQRIKELMAQDMAFIRSQNGHAVD"
7229..8311
/gene="psbA2"
7229..8311
/gene="psbA2"
/note="ORF_ID:slr1311"
/codon_start=1
/transl_table=11
/product="photosystem II D1 protein"
/protein_id="BAAL6586.1"
/db_xref="GI:1651658"
/translation="MTTTLQORRESASLWEQFCQWVTSTNNRIYVGMFGTLMIPTLTA
TTCFIIAIAAPPVDIDIGIRVAGSLLYGNIIISGAVPSSNAIGHFYPIWEAASL
DEMYNGPQVYVHFELIGIFCYMGROWELSLHMRPWCIVAYSAVSAATVFLI
YPTQGSFSGMPLIGISGTENFMTVEQAEHNLHMPHMLGVAGVSGSLFSAMHGLS
VTSLVRETTESVSNQYKFGGEETYNIVAAHGFGRLLIFQYASFNNSRSLHFFLG
AWPVIGITFAMGVSTMAFNLFNGFNQSLDSQGRVIGTWADVLNANLIGFEVWHER
NAHNPDLASGEQAPVALTAPAVNG"
8492..10471
/gene="speA"
8492..10471
/gene="speA"
/note="ORF_ID:slr1312"
/codon_start=1
/transl_table=11
/product="arginine decarboxylase"
/protein_id="BAAL6587.1"
/db_xref="GI:1651659"
/translation="MGEEPVPADKALGKFKKKKNASWISSEALYRVEAWGAPYPAI
NAAGNITVSPNGDRGSLDLLELVEALRQRKGLPLIRFSDILADRLNLSCFKA
IARYNPNTYQAVYVKCQNRHLVAVRFGQTSQGLEAGSKPELMIALATLPPPL
DRODKHTKPLIICNGYKDQYLETALLAKRLGRHPRIIIEQLRELVNHLISQOLNIK
PMLGVARLSQSLKSEISSGNGDRAKGLTMPDIVTVIHLRENNCLDCLKMLHFH
LGTQVSDIALIKEANREASQLYVELYKLRKMYLVNCGGLAVDYDCSKNYPSAKNY
NMQYANDIVAAIQDACLGVSPPIVSESGRAIMAHQSVLFDVLGSLQTSERSEH
PPDENAHPLKLNKWEYETIAEQYQYHDALQKLTAEASLNFGLSLTERGAQO
IHWACCKRIETRLQLEYIPEDQALDQIMTDIYVNLVSFQSAPEWSLDFPLPILP
IHLNKEPVSORVILADITCDSDGKIDRFIDLVKYSLEWHPLENDGNPYLGMFLVG
AYQIMGNLHFLGDIINVHTATTPQGVQIESVVRGDTMTVELGVQVYDSDLLLEGLR
RHTLALNSQGITLESRLLEDYEQSLRRTYLS"
complement(10622..12631)
/gene="lig"
complement(10622..12631)
/gene="lig"

gene
CDS
complement(10622..12631)
/gene="lig"


```
LMVQTGMYTEIGTSSQSLVLRKMGYNLRVTLQNCRLDKAGFNDLVSVNYFNV
IDETLDIOTIAYHRELEAFADKVEPAIFIGLQPHFTHLETYALDKELKPGYDP
MSMPWTAKLWNPEPLGSGFFGVCQAWQPNDFGREVMNILEQRGLKADLETAL
HSPLPDKKFPPTMAEGKLSPI"
complement(1867..2904)
/gene="rpoD"
complement(1867..2904)
/gene="rpoB"
/note="ORF_ID:sl10306"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="BAAL0020.1"
/db_xref="GI:1001398"
/transl_table="MVTVTVILLFIKESFRMPTANLSSPTTADMVRSYLHEI
GRYPLTHQEIOIILGKQVQOQMALEHKKALADRLGRLPSDPEWAAADSVTKLHRY
LGGQERAKRMIEANLVAIAKKYORKNNEFLDLIOEGSLGRLGVERKFDPTKGVK
FSTVAYWIRQATRAAQGRTRIRLPIHITKLNKIKKTQRELSOQLGRSATPAEVA
KALEIDSQREYLSLRSQISLSDVRVGDNDQTELSLLEDEGVSDAYITQESNQD
LQNLALSTLQQQAVLTMRFGLNDGQELSIAKIQHNLISRRVROLENQALQAKRR
RANNAEYIIAS"
3365..4312
/gene="era"
3365..4312
/gene="era"
/note="ORF_ID:slr0321"
/codon_start=1
/transl_table=11
/product="GTP-binding protein Era"
/protein_id="BAAL0021.1"
/db_xref="GI:1001399"
/transl_table="MDIPNTATIAITPQAPGRSFVAIVGRPNVKGSTLMNQLVQ
OKAITSPTAATNRLLQGIITPSSQIILLDPFGHKPHHELGRVLRKNAIQAIHSV
DLVYFLDSTALGRGVRVYDLQKTDGVPVYGLNKDQDQPPDQREELNASEYTLTE
NHGVPKFKSALGGRSFSQSALEARDLDPGVYYPDLYTDQPERFIAMELREOIL
LTLQEPVSHVAIALEKVEETPERTNFAALTVERSGKGIILIQGSMQALGATTAAR
QQTKLSGVYILKLVKVPKVRQSRQQLLEFGYRVEE"
4322..7609
/gene="cheA"
4322..7609
/gene="cheA"
/note="ORF_ID:slr0322"
/codon_start=1
/transl_table=11
/product="chemotaxis protein CheA"
/protein_id="BAAL0022.1"
/db_xref="GI:1001400"
/transl_table="MTSDPNPAQAFENWFEERGGDAFTSLIAGEETNSGIELDFGD
LFEGEPTVIOGEGFSPSGCEOPPVRTLDNSPEALDLAFTDEDAPTAIDPRES
SSRLIEDOMEADPLASIPETNPAYQTAGLDNDLAEQLDGLGVGDEPAINFGL
HSFAGADRVTPLTNEVLFAIDDLGLINAPVDEEFSTFLETLIDTPAPKSFATL
SHFRLPVAEEDDEFKDLAELEQADSSLGIGAGMGGARLPARPPOVKSRSFOT
MRVPIKQDLNLSLIGELVVRNRLQEQRLQFQDLNLLQANLSDVGRMDQDLYE
RTLLEGALLASRNSRHSKTYPDARTASSSTMDQPLDAEMDRFTGPHLLS
QEMTELVRYRESADNLIQVDETDQVARTLRQVTTQLEGMTKSRMVPPTQADRLP
RAIREVSKLNKKNVLEVESGDVLDKMLEHLNPNHLYNNATHGISTPAERQKL
GKSGPEGLHAFQGGQTLITVSDDGAGINTEWKOKAVEKGLISMEASQUSQSEV
YDLDFAGFTKTDADFGAGVGLDVVRTSLDIRGAINIESTVKGKTFITRLPLT
LSICKALCLSDRARIAPMDGVEDTRDYLKDIOTDEQKCVKYEKALLPLLYPL
KDLNLYNRQLSRSGVGGKOEVDVVRVIAAGSSGNLLAQIDQVLGEEIPIVQELG
PIPKPCGIAGTVRGDSIMPIADVLEIEIAOGRLETDSSGGLWRKLTPTVNEVHR
DPVLYVDDSDITVRELLSLSFKKAGFRVQARQOEAMQDKSLGCLPDLVFCDIEMPR
KNGLELLEQADDELSIPVALLTSRGADRRKRAAKLGATGFTFKPTKEDILDDAA
NRMNGVLLPGSTKRTNLSLRFKKNRNFSEAPITVRSNFTGLSTVLIIDDSVA
VREMLSMFTFKKADYHTARDQDAMEKLSGLPCDLICDIEPMRNGLELLERIOE
DPNLSRTPVAMITSRGAQKQRMARAAEKANGKYFKVPIETTLLEASQRLIRGEVLDL
AALSAD"
7609..10737
/gene="amsI"
7609..10737
/gene="amsI"
/note="ORF_ID:slr0323"
/codon_start=1
/transl_table=11
/product="alpha-mannosidase"
/protein_id="BAAL0023.1"
/db_xref="GI:1001401"
/transl_table="MTESLSPILLPEISIALVLEQLRRLCRONCQSHWYCLPGAPWEIS
DNQVTPLGKVNKGYLWEKSEKTOFACQTFVVPKALADYPLDGLTLRLALTWAKDA
QIFVNGELVAGLDSDKARVLLTNSQIGRTFTVALKTLSSGHDIGGLQMSQLIFER
KYPDLDPGFIADEIEVLSLYLAQFESERLAQSKLTLEATWNKISDQEQFDSLQLR
ORLLLTLEGLKQOQNLGLGHAHDMALWQLEETWEVGERTFQSVINLOTEFADLVFG
HTSPVLYOIEEHPALFOIOVAIKGNWELLGGMWVEPVNIISGESLARQFLYGO
RYFEAKFGOISKVAWLPDSFGGOLPOIFHOSGIDYFVTGKLHNDTNSFPBGAFHW
RSPDGTETVMSPPNLAGVMDNQPLPMANYANRWQEQGLKECLWLPVGVGGGGPS
RDMWENKRNQSEFFPTINTAKAEHLAGIKQKQSDQSFPIWERELYLELHRCGFT
VHADQKLYNRQCEHLYEAEFLWSFANWLGYSYPRKELAWKLLNQFDLPFGT
SPEVEVTANQAEVYQNTNIIQAALKHLSQINPSYHSDSTERETPLLVNPLKH
WQSOQVLEPIQOEKHYQVYDQGTTLVNSQITAEGLIFLVDSPPLTKYWLKTKKE
PLNSQNLDSFSQDPTLNLGLLKVESLQNGEIASYDLANQROVLGAGHNOLOQFRD
QGOYDANDNDPNVSOYPLSSAKLNYORLEQGPLHWRVVKKQSQSVETQDYCLTK
DSPLLEIKTEVDQWETHVAKAFPIALQSEFTTEAPCAVVDRTNPTNTEKAKQME
VPHQWFSLTDQNTENYVSLNKNKYGCDVKNLMLRLTLRSVSWPDPDSADKGNKFS
YTFPHQGNQAKTAVAGYTHRPLRTYFENFNPPLSSQKLLSPCEWLNLSGDNLC
LMALKLREYDPQWVLRCEYMSGESAQLEVITAFDRHVKTRLNLLLEAEQPEENNHNH
PMQIVSVCLAYD"
10730..11614
/gene="appC"
10730..11614
/gene="appC"
/note="ORF_ID:slr0324"
/codon_start=1
/transl_table=11
/product="oligopeptide transport system permease protein"
/protein_id="BAAL0024.1"
/db_xref="GI:1001402"
/transl_table="MTKTPPLSQSVAINLQKRFQKPLASLSLITLIIVLSVWFGP
IFVRPIDQIDFSQTAVPSPWOPLGTNDLQDOLARLLIGRISLAVGLSAMAVSLT
LFTLGAAGYGGWMDVILMRITDLFLALPOLPLVLLVYLFERSEVTRATSPKGIIF
ILVVLIGLUNMSVARNLVRGNILRLRENEFVKAALAMGAGSNHIIWRHLLNVLNLI
IVAASLVAGNAIITESTLSLGLGFPDPVPTWGMQMLYAKDYLETAPLMAVFPGLAIF
ITVLSVNYLGDGIGIKNS"
11662..12147
/note="ORF_ID:slr0325"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAAL0025.1"
/db_xref="GI:1001403"
/transl_table="MIRPVPDASNOFYLDHINLLRQSYERLTGKTGPGSLGVALAQ
AVEAPVYVVSIGIEPDPPIFNANLTAQKLEFPWVEFCQLPSRQSAELPNQTERQOL
```

alignment_scores:

Quality:	298.00	Length:	187
Ratio:	2.614	Gaps:	3
Percent Similarity:	60.963	Percent Identity:	38.503

alignment_block:
US-09-252-691-7056 x SYCSLRA ..

Align seg 1/1 to: SYCSLRA from: 1 to: 125469

44	IleLeuPheAsnLysProTyAspValLeuProGlnPheThrAspGluAl	60
78736	TTATGGTTAATAAACCGAGGATGCTGCTGCCACCTGTGACGATCCCG	78785
60	aglyArgSerThrLeuLysAspPheProVal.....G	72
78786	GGCGCGAAACAGTGCCTGGATTTACTGCCCCAGGACTTACAGCGGGCA	78835
72	lnGlyValTyAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeu	88
78836	AAGGACTCATCCCGTGGCGGTTAGACCGTAATTCACCTGCGGCTCTC	78885
89	ValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLy	105


```
78886 CTGCTCACCATTGATGGCGAATGACTCTCGCGGTGACCCATCCCGTTA 78935
105 sargThrGlyLysIleTyrValGlnValGluGluProAspAspA 122
    :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78936 CCATTGGCAAAACCTATGACGTTTGGTTAGAGGTAATCCAGTGATG 78985
122 laSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr 138
    :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78986 AAGATCTGGAAGATGGCGATCGGCGATGATGATGATGATGATGATGATG 79035
139 LeuProAlaGlyIleGluArgValAsnGluProGluThrLeuThrProAs 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79036 CTACAGCTACTCTGGAAGTAATTCGAAATAAA..... 79071
155 gAsnProProIleArgGluArgLysSerIleProThrSerTrpLeuLysI 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79072 ..... 79090
172 leThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHis 188
    :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79091 TGACGTTGACGGAGGAGCGTAAATCGCAAAATTCGCTTTAGCGGAAGAA 79140
189 ValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrTh 205
    :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79141 TTGGATTGACGGTCTGAAGCTCATCGCCGACCATGCGCCCTTACA 79190
205 rLeuAspSer.....LeuAlaAsnGlyGluTrpArgAspValt 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79191 GTTGATACACGAGGCGCAAGTCTCGGTAGTGCGCAATTCGTTTCITT 79240
218 hrProLysGlu 221
79241 CCCCGCGGAA 79251
seq_name: gb_ba2:AE001942
seq_documentation_block:
LOCUS AE001942 9999 bp DNA BCT 22-NOV-1999
DEFINITION Deinococcus radiodurans R1 section 79 of 229 of the complete
            chromosome 1.
ACCESSION AE001942 AE000513
VERSION AE001942.1 GI:6458611
KEYWORDS
SOURCE Deinococcus radiodurans.
ORGANISM Deinococcus radiodurans.
REFERENCE 1 (bases 1 to 9999)
AUTHORS Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
Smith,H.O., Venter,J.C. and Fraser,C.M.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1
JOURNAL Science 286 (5444), 1571-1577 (1999)
MEDLINE 20036896
REFERENCE 2 (bases 1 to 9999)
AUTHORS White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
Smith,H.O., Venter,J.C. and Fraser,C.M.
Direct Submission
Submitted (08-Nov-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source 1..9999
            /organism="Deinococcus radiodurans"
            /strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
/complement(71..655)
/gene="DR0891"
/complement(71..655)
/gene="DR0891"
/note="similar to GP:1498492 percent identity: 61.93;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-binding response regulator"
/protein_id="AAF10469.1"
/db_xref="GI:6458612"
/translation="MRLVIADHPFLRMGLKYALIQOQFVVVAEADGSPALDACTY
QPDAAALDKVPMGTGIEVCDRLRTAPRVSVILITFAEPAIVQAAARAAGAKGVSK
ETDPESLARQLREIIVAHVEVDRLPOVDVPRPLTPRESEVLPLLRGFSNKEIARDLGS
PDTVKDLARLYSKLNAGDRTEAVSRARSIGLS"
/complement(652..1761)
/gene="DR0892"
/complement(652..1761)
/gene="DR0892"
/note="similar to PID:2245589 percent identity: 47.14;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAF10470.1"
/db_xref="GI:6458613"
/translation="MTRTPPAAPRAAAAPPERRPARLAAVAGSRPPLOTQTQFT
LVSLIAPLNVLALASGAGSLSSAAWMLVGLLSAVIGYVLSGALLRLRLLEL
EVSDEGFAQHADDREINSLRRAFSGSLMRLTGTEQTRNFMATLVHDLKTLPLIATG
HLTRALTEPLPDEKRAVQEIQTETERLLAQVQOMADHFRDDVQLHLPAGDLR
EVAQAAARLRPQAEARGLSHVGSGQAEIDAAALERAVALGNLTNALYARHEVRLT
AAPGLLRVDDGPGLSASLDLAQPFNAQVIAQOQYTAGTAGLGLFIVRRIAEAHG
GTLRYGREPPRGRLPAAPCAFELTTLVLPEVLP"
/complement(1922..2617)
/gene="DR0893"
/complement(1922..2617)
/gene="DR0893"
/note="similar to GB:U00096 PID:1787004 percent identity:
66.98; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF10471.1"
/db_xref="GI:6458614"
/translation="MVKSMQOIAMTQOKTLDQVTRFMARTYSMAAGLALTAGVAVLT
AQNEGLAQVASLRPLMLAQALVFLSMFAQLRISAAGALFVGYAALTGLTFSAL
LFAYSPAAVITAFVAGTGLMSVAGFVKKDLKSAMGRFFLFAVLGLVAMLVNFV
GSSLSLGSIMIGVFLFAGLTAYDTQMLRNALSGISGEQASRASINGALALYLDFIN
IFLFLNLINGSRD"
/complement(2702..4207)
/gene="DR0894"
/complement(2702..4207)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF10477.1"
/db_xref="GI:6458620"
/translation="MIKKEILGRPVAVSNGAQVDNVHDVDFHOGNRVALLVEEG
GWLRAKAVPFRGVQSIGEKAVMTAPEDVNTGDPVLKEALAGTNLIGMTLLTDD
GTLCKITDVFEDERSGVEGVEASGGLFADMTNGRTFFPAQSIQIGKDSALVPAGV
AAAREQAPGGIRAFHAGTACKSLGNAVGAASRGVAYEDTSRRENYEQVSANVAE
AAREQQQARAASADLSRVSQGAEDLEASEROKDFVGVKAGVDVTDNGVLVA
KCTTIDEGQAREQNGALPALLAAAGSREVRDLADPSAGADVYVGRVQEDV
RSSVGLSLAVQGOIVTPEVARERLEERLVAATQAPLOTPPVPPRAELREQASD
LRNLADGLSNYSOVNOCANELIGRAKSWLGERREDLEDLSLDROEOEAQARIDAL
GRPVTRAVLTTRDRLVILQPGELVTHALDEARAAGVLGLLLSDVGGGAPLS"
4311..4697
/gene="DR0895"
4311..4697
```


211 scNlyGluTrpArgaspValThrProLysGlu 221
| | : | : | : | : | : | : | : | :
5333 CCGCGGAATACGGAGCACTCCGCCGCCTGAC 5364

sseq_name: gb_ba2:AE001708

sseq_documentation_block:
LOCUS AE001708 19531 bp DNA BCT 02-JUN-1999
DEFINITION Thermotoga maritima section 20 of 136 of the complete genome.
ACCESSION AE001708 AE000512
VERSION AE001708.1 GI:4980740
KEYWORDS
SOURCE Thermotoga maritima.
ORGANISM Thermotoga maritima
REFERENCE Bacteria; Thermotogales; Thermotoga.
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Utterback,T.R., Malek,J.A., Linner,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.

TITLE Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima
JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316
FEERENCE
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Utterback,T.R., Malek,J.A., Linner,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.

TITLE Direct Submission
JOURNAL Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1..19531
organism="Thermotoga maritima"
db_xref="taxon:2336"
104..226
/gene="TM0242"
104..226
/note="TM0242"
/note="similar to percent identity: 0.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
protein_id="AAD35333.1"
/db_xref="GI:4980741"
translation="MRLEAKQSIPEFLNGFSEISGVSRVRLRQREYEFEVEK"
223..1074
/gene="TM0243"
223..1074
/note="TM0243"
/note="similar to GB:Pyr.h percent identity: 58.74;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
protein_id="AAD35334.1"
/db_xref="GI:4980742"

SSTPSHLSRVTKWEVEDLLGLNDPRFKVCIVYSPSYSGKDRLIIPRPVSYSVRASIEEVEVFELGADRGVIADVFNKLFEKGRIERHLHLKEKVSMPFGRITTVIGLGESKDIVDFTWVARERNIVSLFAFTPDKGTAFENRERSLEYRKIQLVTYLKLNKKPNIIFDSNGKIIDVENGEFFEEALTRGCPHCPTRYNESRPGIINYVHR"
1031..2386
/gene="TM0244"
1031..2386
/note="TM0244"

gene 1031..2386
CDS 1031..2386
/note="similar to PID:893416PID:893415 percent identity: 100.00; identical by sequence similarity; putative"

[illegible]


```
alignment_scores:      Quality: 253.00
```

complement(996. .1649)

gene


```

/feature="ypfa"
complement(996..1649)
/feature="ypfa"
/function="unknown"
/feature="alternate gene name: jofa"
/codon_start=1
/transl_table=11
/protein_id="CAB14223.1"
/db_xref="GI:2634726"
/db_xref="SWISS-PROT:P38491"
translation="MIEIGENVLLYIEBENELKKAASKAVSIENNELLIAVPDVVTVG
RTVILHNDMEVTEVGVDEVPYRISRIKGVKDKLQMLCEMPPREKMKRIORROY
VTDAYLVQIPGNEBIRLISYINISAGGIYAVLADLSFGSGSLIRLIRLPEEH
TQIQTAEVVRIFNDPSEKRRMTLEYSEIAGGQALLQICIRQLNKRKARME"
complement(1711..3094)
/feature="ypeb"
complement(1711..1733)
/feature="ypeb"
complement(1742..3094)
/feature="ypeb"
/function="unknown"
/feature="alternate gene name: joeb, yzuA; similar to
hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14224.1"
/db_xref="GI:2634727"
/db_xref="SWISS-PROT:P38490"
translation="MIRGILIAVLGIAVGTGYGKHEQEKDAVLLHAENNYQRAFH
ELTYOVDLQDKIGTILAMNSQKSLPDLIDVWRTSEAHNSVSLPLTLPFNKTEE
LLSKGDSYKTSYRDLQKPDKNREYSLNKLQYSEDQINELRHVHLVMSKLNLRW
MDVEVALASDEKQAPDNYISIKVYSGKKTNRDYSISMKDPDHPKAVIYMDITKKG
QISEQAKQIAERFADPNYSIKVYSGKKTNRDYSISMKDPDHPKAVIYMDITKKG
HEVYIQREAVKQRIISNDGNALAFKNGFETDLEIDESAQYDKIGVFSYVPV
ENKRYMPEAIRKVALDQGVVGSFARDFLTSKRITPKPAITEAKSKLNKQVQ
VRETRALITNELGOELCYEMLTGTIENDTFBMYINAKDGSSEKVEKLNAPPIYKDL"
complement(3129..4046)
/feature="sleb"
complement(3129..4046)
/feature="sleb"
/feature="alternate gene name: ypea"
/codon_start=1
/transl_table=11
/product="spore cortex-lytic enzyme"
/protein_id="CAB14225.1"
/db_xref="GI:2634728"
/db_xref="SWISS-PROT:P50739"
translation="WKSQGSINACILILSFSTITTTINTETISAFNSQVIOGATGDDV
VELQARLQNGYNGKIDGVGWGTYWAVNFQDQGLKEVDGLGAKTKQTLLCKSK
YYREYVMEQLNKGNTFTHYGKIPKQYKPSKAATQKARQQAERKQKQPAETKQPK
ANAKQOQNPAPAKRQDAVAANMPGFSNNDIRLLAQAVYGEARGEPEYEGQVATAAV
ILNRLNSPLFNSVAGVIFPEPLAFTAVADQIYMQPNETAREAVLDAINGWDPSEAL
YFENPDATSPKMGWRPQIKRIGKHICE"
complement(4162..4841)
/feature="ypdc"
complement(4162..4182)
/feature="ypdc"
complement(4185..4841)
/feature="ypdc"
/function="unknown"
/feature="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14226.1"
/db_xref="GI:2634729"
/db_xref="SWISS-PROT:P50738"
translation="MFAISAGIAPGIALLSYFKQDQVNEPVMVLRSLFGLGVVLV
FFIMTQYVLEKENVGGGFFYSFSSGFLSESLKWFILMISVYPAHFDEHYDGIVY
GASVSLGFTALNLYLIGHGHAFAVRALLPVSCHALIGVIMGYLGLKARFSADKAR
VKWLTLSLVVPSLLHGSYDFILTALSNWYIYMLPFVFLWFLWFLKAKKARSVMQV"
complement(4929..5935)

```

```

/feature="ypda"
complement(4929..4953)
/feature="ypda"
complement(4961..5935)
/feature="ypda"
/function="unknown"
/feature="similar to thioredoxin reductase"
/codon_start=1
/transl_table=11
/protein_id="CAB14227.1"
/db_xref="GI:2634730"
/db_xref="SWISS-PROT:P50736"
translation="MIQKAILIGGCGCLSNAAHLKQIGIDALVIEKGNVNVSYNY
PTHQFFSSSEKLEIGDAFITEENKPVRIQALSYREVVRKKNIRVNAFENVRKVT
TONFTVLETSEKTYTTCYCIATGYDHPNMGVGEDLPKVFYKFGHGFYFKDV
VVIQKNSVDAALELVKSGARVTVLYRGNYSPIKWPILPEFALVRNGTIRMEFG
ACVEKITENVEVFRSGEKLITIKNDFVAMTGYHPDHFQLEKIGVEIDKETGRPFEN
EETMETNVEGVFIAGVIAAGNANEIIFIENGREFHGHGHAETAKRENH"
complement(5995..7324)
/feature="ypca"
complement(5995..6016)
/feature="ypca"
complement(6044..7324)
/feature="ypca"
/function="unknown"
/feature="similar to glutamate dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB14228.1"
/db_xref="GI:2634731"
/db_xref="SWISS-PROT:P50735"
translation="MAADNRTGHEEDKDLVKLTQTVIHKALEKLGYPEEVELLKE
PMRLTVIPYRMDGSGYKFTGIRAHNDSYGTGKGGIRFHPNVTKEVKAVKALSIW
MSLGGIIDLPGYGGKGVCDPRDMSPRELERISRGYRAISQVGTQKVPAPDVF
TNSQIMAMWDEYSRIDFENSPGFTGKPLVLGGSHGRESATAGVTICKEAAKRG
IDIKGARVVGVFGNAGSYLAKFMHDAGAKVVGISDAVGGLYDPESGLDIDYLLDRDS"
alignment_scores:
Quality: 253.00 Length: 182
Ratio: 2.259 Gaps: 5
Percent Similarity: 61.538 Percent Identity: 35.165
alignment_block:
US-09-252-691-7056 x BSUB0013/rev ..
Align seg 1/1 to reverse of: BSUB0013 from: 1 to: 218470
39 GlnProThrArgValIleLeuPheAsnLysProTyAspValLeuProG1 55
25969 GAGCGGGTTTACTTCTCTATAT...AAACCAAGAGCGGTATTCGCG 25923
55 nPheThrAspGluAlaGlyArgSerThrLeuLysAspPhe.....I 69
25922 AGCCAGATGATAAGAGCGCAAGTGGTGACGGACTTTTAAAAATA 25873
69 leProValGlnGlyValTyAlaAlaGlyArgLeuAspArgSerGlu 85
25872 TTCGG...CAGCGCATTTATCCGATTGGCGTCTGGACTATGATCAGC 25826
86 GlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrG1 102
25825 GGTCTTTTGTCTTTTAACTAATGACGGCAATTCGCCAATAGCTAATGCA 25776
102 nProGlyLysArgThrGlyLysIleTyTyValGlnValGluGlyGluP 119
25775 TCCTAAGTATGAATAGACAAACATACGTGGCGGAGGTGAAGGAATCC 25726
119 roAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAsp 135
25725 CGCCTAAGAGCTCTCAGAAAGCTGGCGCGCAATAGCTGAGGAA 25676
136 GlyProThrLeuProAlaGlyIleGluArgValAsnGluProGluTrpLe 152

```



```

||||| ..... ||| ||| ..... |||
276 AGCTACGTAGCAAAATTAAGGTGTTCCCACTAAACGGGATCTATTCG 325
126 ysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGly 142
||| ..... ||| ||| ..... |||
326 CATTAGCCAAAGGCATTCGCAATGATGCAACAGCAGCAGCCACCAAT 375
143 IleGluArgValAsnGluProGluTrpLeuTrpProArgAsnProIle 159
||||| ..... ||| ..... |||
376 TTTCAATTCCT ..... TCAGCGGACAT 398
159 eArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTrp 176
||||| ..... ||| ..... |||
399 TAAACAGGTAGCAGTGTC ..... GTGGAATTAACGATTCATG 436
176 lucLeuArgAsnArgGlnValArgMetThrAlaHisValGlyPhePro 192
||||| ..... ||| ..... |||
437 AAGGACGTAAATCATCAAGTGAAGAATGTTTCAAGCTGTGTTCCCT 486
193 ThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeuAspSer 209
||||| ..... ||| ..... |||
487 GTTCAAAATTAACAGCTGACGTTACGTTGAGTTGACCTTCAAGGTT 536
209 uAlaAsnGlyLutTrpArgAspValThrProLysGlu 221
||||| ..... ||| ..... |||
537 GCGACAGGCAATATCGTGTTTAACTAAATAAAGAA 573

```

seq_name: N_Geneseq_36:V52297

seq_documentation_block:

```

ID V52297 standard; DNA; 9707 BP.
AC V52297;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:164.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN W09818931-A2.
PD 07-MAY-1998.
31-OCT-1997; U19588.
31-OCT-1996; US-029960.
(HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 1057-1063; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridize to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 9707 BP; 2721 A; 1959 C; 2369 G; 2658 T;

```

alignment_scores:

Quality: 246.50 Length: 179
Ratio: 2.221 Gaps: 3
Percent Similarity: 62.011 Percent Identity: 33.520

alignment_block:

US-09-252-691-7056 x V52297 ..

Align seg 1/1 to: V52297 from: 1 to: 9707

```

45 LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGly 61
||||| ..... ||| ..... |||
5010 CTGCTTAACAAACCCACGCGGTGTCATTTCACGTGTGACAGATGATAAGG 5059
61 YArgSerThrLeuLysAspPheIlePro ..... ValGlnGlyValTyrA 76
||||| ..... ||| ..... |||
5060 TCGCAAGACGGTTGTCGACCTCTTGGCCCAATGTCAAAGAGCGTATTTACC 5109
76 laAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThrAsn 92
||||| ..... ||| ..... |||
5110 CTGCGGCTGTTTGGACTGGGATACATCAGGTGCTTGATTTCACCAAT 5159
93 AspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGlyLy 109
||||| ..... ||| ..... |||
5160 GATGGGACTTTACAGACGAGATGATTCACCTCGTAAATGAGATTGACAA 5209
109 sIleTyrTyrValGlnValGluGluProAspAspAlaSerLeuAlaL 126
||||| ..... ||| ..... |||
5210 GGTATTATGTCGCGGTGTTAAAGGTGGCCCAATAAGGACAATCTCGCC 5259
126 ysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGly 142
||||| ..... ||| ..... |||
5260 CTTGACCGCGTGTCTGAGATTGATGTAAGAAACCAACGACGAGTGT 5309
143 IleGluArgValAsnGluProGluTrpLeuTrpProArgAsnProProI 159
||||| ..... ||| ..... |||
5310 TATGAATTCCTC ..... AAAGTGGACCCAGT 5335
159 eArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTrp 176
||||| ..... ||| ..... |||
5336 CAAAAATCGC ..... TCTGGTGCAGTTGACATTCATG 5370
176 luGlyArgAsnArgGlnValArgMetThrAlaHisValGlyPhePro 192
||||| ..... ||| ..... |||
5371 AAGGCGGTAAACCATCAGGTTAAAAAGATGTTGAAGCTGTTGGTCTCAA 5420
193 ThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeuAspSerLe 209
||||| ..... ||| ..... |||
5421 GTAGATAAGTTGTCGCGACTCGTTTCGGACACCTAGACTTGCACGAGCT 5470
209 uAlaAsnGlyLutTrpArgAspValThrProLysGlu 221
||||| ..... ||| ..... |||
5471 CCGTCCAGGAATCCCGTCTCTTAATAAAGAA 5507

```

seq_name: N_Geneseq_36:V65255

seq_documentation_block:

```

ID V65255 standard; DNA; 5816 BP.
AC V65255;
DT 24-DEC-1998 (first entry)
DE DNA encoding S. pneumoniae proteins of unknown functions.
KW Streptococcus pneumoniae protein; recombinant; gene expression;
KW DNA chip; virulence; antibody; infection; detection; treatment; ss.
OS Streptococcus pneumoniae.
PN W09826072-A1.
PD 18-JUN-1998.
PF 09-DEC-1997; U25278.
PR 13-DEC-1996; US-036281.
PA (ELIL ) LILLY & CO ELI.
PI Baltz RH, Bargett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI Mills BJ, Norris FH, Peery RB, Rosteky PK, Rosteky PR,

```



```

536 ATGTTGCAATAACACACAGGTTGTCATT...TCTGGACGGGAAGATCAGCG 490
61 yArserThrLeuLysAspPheIleProValGlnGlyValTyrAla.... 76
   :|||||: |||:|||||: |||
489 AGATGAACGGTAATTGACTGTTTAACGTATCAGGATMCCGTGCGGACT 440
77 .....AlaGlyArgPheAspArgSerGluGlyLeuValLeu 90
   :|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:
439 TGTGCCCGTGGTGGCTGGATGAAGTTCGCGAGGCTTCTCTACCA 390
91 ThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgTh 107
   :|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:
389 ACAATGATGCCAGTATGACCAATCAATATATCMCCCAAAAGCATGT 340
107 rGlyLysIleTyrValGlnValGlnGly.....GluProAspAspA 122
   : ||| ||| :|||:|||||: |||:|||||: |||:|||||: |||:
339 GGATAAARGAWTATTAGCTGAAGTTGAAGGACTTGTGACAGCGGAAGATT 290
122 laSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr 138
   :|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:
289 GTCAACTG.....TTCCGCAAGGATTTCCTTGAAGATGCGAAGATG 246
139 LeuProAlaGly.....IleGluArgValAsnGluProGluTyrPle 152
   :||| :||| :|||:|||||: |||:|||||: |||:|||||: |||:
245 GTTAACCAAGTCGTTAGTCATTGAGACGATCAATGAGGAAGA... 201
152 utrProArgAsnProPheIleArgGluArgLysSerIleProThrSert 169
200 .....GGGACTTCGG 191
169 rpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgArgMet 185
190 TAGTCCGTTTATGTTCAAGAGGGGAAATTCATCAAGTCAACAGCGATG 141
186 ThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetG1 202
   ||| ||||| |||: |||: |||: |||: |||: |||: |||: |||: |||:
140 TTAGCCACTGTAGGCAACCTGTGACGTATTATAACGATTACGAATGGG 91
202 ySerTyrThrLeuAsp....SerLeuAlaAsnGlyGluTyrPArgAspValt 218
   :||| :||| :|||:|||||: |||:|||||: |||:|||||: |||:
90 CGAACTAGTGGTGGATGAACGTTAGACTAGGGGAATATCGGCCCTTAA 41
218 hrProLysGlu 221
   |||
40 CAGAAGCCGAA 30

```

seq_name: N_Geneseq_36:V74878

```

seq_documentation_block:
ID V74878 standard; DNA; 676 BP.
AC V74878;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #567.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
  cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
  skin infection; surgical wound infection; scalded skin syndrome;
  toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 421..480
   /tag= a
   /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PF -07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA,
DR WFL; 97-374922/35.

```

```

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 1505-1506; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
of the invention. The DNA sequences are recorded on a computer readable
medium, preferably selected from a floppy or hard disk, random access
memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
the S.aureus DNA sequences allows putative functions to be assigned so
that protein-encoding or regulatory regions of commercial, therapeutic or
industrial importance can be obtained. Specifically, sequences which are
likely to encode antigens have been identified and these polypeptides can
be used in a vaccine composition against S.aureus infection. The
polypeptides can also be used in a kit for the immunodetection of
S.aureus in a sample. S.aureus is implicated in numerous human diseases,
including cellulitis, eyelid infections, food poisoning, osteomyelitis,
skin and surgical wound infections, scalded skin syndrome, toxic shock
syndrome, etc. Organisms transformed with the DNA sequences can be used
for recombinant production of the polypeptides. The new DNA sequences
and their fragments are useful as primers or probes for isolating
homologues of any of the S.aureus DNA sequences contained on the
computer readable medium.
SQ Sequence 676 BP; 163 A; 136 C; 77 G; 236 T;

```

alignment_scores:

```

Quality: 201.50      Length: 180
Ratio: 2.056         Gaps: 2
Percent Similarity: 54.444 Percent Identity: 31.111

```

alignment_block:

US-09-252-691-7056 x V74878/rev ..

Align seg 1/1 to reverse of: V74878 from: 1 to: 676

```

44 IleLeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAl 60
   ||| |||||: ||| :|||||: ||| :|||||: ||| :|||||: ||| :|||||: |||
537 ATTTTTCCTAATAACCAACTCAAGTGTATMACAAGTGTATCTGATGATAG 488
60 aGlyArg.....SerThrLeuLysAspPheIleProValGlnGlyValt 75
   ||||| :||| :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||
487 AGGAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 438
75 yrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThr 91
   :||| :||| :|||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:
437 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 388
92 AsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrG1 108
   ||||| ||| :||| :|||: |||: |||: |||: |||: |||: |||: |||: |||
387 AATGATGTTGAATTTACTAATTTAATGACACATCCCAAGATATCAAAATTA 338
108 yLysIleTyrTyrValGlnValGluGlyGluProAspAspAlaSerLeuA 125
   ||| ||| :||| :|||: |||: |||: |||: |||: |||: |||: |||: |||
337 GAAAAATATGTTCCGAAATTAAGAGTTATTATTAATGAGAGAAAGATGA 288
125 laLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAla 141
   :|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
287 AAGCGCTAGAAAAAGGTATTGAATTAGAAAGTGTATGACGCAACCGGCT 238
142 GlyIleGluArgValAsnGluProGluTyrPleThrProArgAsnProPr 158
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
237 GAAGTGAATA.....GT 227
158 oIleArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeut 175
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
226 TAAGAAACAAGATAAAGATAAAACACACATTTGTTGAAATTTACAATTA 177
175 yrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhe 191
   :|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
176 CAGAAGGTGCTAATCGCAAGTCAAGAGAAATGTTTCGAACATTTTGGTCAT 127
192 ProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeuAspSe 208
   :||| :||| :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||

```

126 CAAGTGACGAACTGCTCGTATTGAATATGGCCCTTTAAATGTTGTCGG 77

208 rleuAlaAsnGlyGluTrpArgAspValThrProLysGlu 221

76 TTTGAATGCTGGTGAAGGACGCTGATTGACGCTCATGAA 37

seq_name: N_Geneseq_36:V52353

seq_documentation_block:

ID V52353 standard; DNA; 4692 BP.

AC V52353;

23-OCT-1998 (first entry)

Streptococcus pneumoniae genome fragment SEQ ID NO:220.

Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

PN WO9818931-A2.

PF 07-MAY-1998.

PR 30-OCT-1997; U19588.

PI (HUMAN) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,

PI Kunsch CA, Rosen CA;

DR WPI; 98-27225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

PS Claim 1; Page 1230-1232; 1409pp; English.

CC The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded

CC on it, or a representative fragment or a sequence at least 95% identical

CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1

CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus

CC pneumoniae. The present invention also describes an isolated nucleic acid

CC molecule encoding a homologue of any of the fragments of the S. pneumoniae

CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced

CC by a process comprising: (a) screening a genomic DNA library using as a

CC probe a target sequence defined by any of the sequences in SEQ ID NO:1

CC to 391, identifying members of the library which contain sequences

CC that hybridize to the target sequence and isolating the nucleic acid

CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced

CC from an organism, amplifying nucleic acid molecules whose nucleotide

CC sequence is homologous to amplification primers derived from the

CC fragment of the S. pneumoniae genome to prime the amplification and

CC isolating the amplified sequences. The computer readable medium can be

CC used in a computer-based system for identifying fragments of the

CC S. pneumoniae genome of commercial importance, or expression modulating

CC fragments of the S. pneumoniae genome. Products from the present

CC invention can be used in diagnosis kits and assays, and pharmaceutical

CC compositions and vaccines for S. pneumoniae.

SO Sequence 4692 BP; 1429 A; 1008 C; 820 G; 1434 T;

alignment_scores:

Quality: 185.50

Ratio: 1.718

Percent Similarity: 57.143

Percent Identity: 30.688

alignment_block:

US-09-252-691-7056 x V52353

Align seq 1/1 to: V52353 from: 1 to: 4692

39 GluProThrArgValIleLeuPheAsnLysProTyrAspValLeuProG1 55

3072 AAACCTCTGTCGCGTACAGCAACAAACAGGAA...CTTCG... 3116

55 nPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIlePro... 70

3117ACCGTCATGGACCTGCTCCATCTA 3141

71 ..ValGlnGly.....ValTyrAlaAlaGlyArgLeuAspArgAspSer 84

3142 ACATCCAGCTGACAAAGCTCTATCGGTGGCCGACTGGACCGAGATACA 3191

85 GluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuTh 101

3192 ACGGAGCTCCTCCTTGAACGATACGAGTCCCTTGGGCTTTCAGCTCCT 3241

101 rGlnProGlyLysArgThrGlyLysIleTyrValGlnValGluGlyG 118

3242 CCATCCCAATATCATGTCGATAGAGCTTACCAAGTTGAGGTTAATGAC 3291

118 luProAspAlaSerLeuAlaLysLeuArgAsnGlyValThr...Leu 133

3292 TTCTAACACCTGACCATATCCAAACCTTTCAAAAGGAATGTCTTTTA 3341

134 AsnAspGlyProThrLeuProAlaGlyIleGluArgValAsnGluProG1 150

3342 GATGACACTGCTGTAAACCCGCAAACTAGAGATTCTATCT..... 3393

150 utrLeuTrpProArgAsnProProIleArgGluArgLysSerIleProT 167

3394GCAAGTCTCTCCC 3396

167 hrSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArg 183

3397 TCAGTCAAGCCTCTATCACCATTTCAGAAAGGAAATTTTCATCAATCAAG 3446

184 ArgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAl 200

3447 AAATGTTCTCTCGGTGGTGAAGTGAAGTCTAGCTCAAAAGAAATCCA 3496

200 aMetGlySerTyrThrLeuAsp...SerLeuAlaAsnGlyGluTrpArgA 216

3497 ATTTCGGGACITTCACATTGAACCCAGATTAGCAGAGGTAACCTACCGCC 3546

216 spValThrProLysGlu 221

3547 CTTTGAACCAAAAGAG 3563

seq_name: N_Geneseq_36:V52508

seq_documentation_block:

ID V52508 standard; DNA; 720 BP.

AC V52508;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:375.

DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

DE computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

PN WO9818931-A2.

PF 07-MAY-1998.

PR 30-OCT-1997; U19588.

PI (HUMAN) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,

PI Kunsch CA, Rosen CA;

DR WPI; 98-27225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

PS Claim 1; Page 1388; 1409pp; English.

CC The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded

CC on it, or a representative fragment or a sequence at least 95% identical

CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1

CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus

CC pneumoniae. The present invention also describes an isolated nucleic acid

CC molecule encoding a homologue of any of the fragments of the S. pneumoniae

CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced

CC by a process comprising: (a) screening a genomic DNA library using as a

CC probe a target sequence defined by any of the sequences in SEQ ID NO:1

CC to 391, identifying members of the library which contain sequences

```
Align seg 1/1 to: XZ0248_07 from: 1 to: 110000
      2 IleMetArgGlnLeuIleThrProGluAsnThrMethrLysThrSerPh 18
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
85250 ATAGAAGAAACTTGTAGAGGTAATAACACATATTGCTAAGCTGGGA 85299
      18 eaRG.....LyshIsArGvaIGluArGP 26
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
85300 TAAGTTACTTTTAGGACAGACAATAATTTATAAAAACAGATTTTTCTTT 85349
      26 heSerSerArgGlnAlaThrArcArgThrProGluProGlnProThrArg 42
```

43 ValIleLeuPheAsnLysProTyrAspValLeuProGlnPhetHraspG1 59
85350 IIAAGAGALLIICAGALLIAGALLIAGAGALLI... 85359

```

59 uAlaGlyArgSerThrLeuLysAspPheIle...ProVal.....GlnG 73
      |||||:::      :::::  |||:::  :::

```

85425 TGA TGG AAG AAG TTAG CAA TAT A TCT TGG T TCA G C C C T TAT TTA AAG AGC 85478

[illegible]

03742 I T T A G A C G A A G G G G A W A A A A T A S A G A A A T A M G A T A A A G G T I I I I I I I G A G C A A A G
190 GlypheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrrThrLe 206
 : ::::: : ::::: ||| : ::||| :::: : ::|||

```

85772  AATATTTTAAAAAATTCATAGAAATTAGAAATGCGCAATATATTT 85821
      206 uasPserLeuAalsnGlyClu 213
      |||||
85822  AGATGTTTTAAAGGAGGTCAA 85843
      |||||
seq_name: N_Geneseq_36:T84021
seq_documentation_block:
ID T84021 standard; DNA; 584 BP.
AC T84021:

```

DNA encoding a Staphylococcus aureus protein of unknown function
 Staphylococcus aureus protein; ribozyme; antisenase sequence;
 Staphylococcus aureus protein; ribozyme; antisenase sequence;
 Staphylococcal gene; regulatory element; bacterial gene expression
 vaccine; Staphylococcal infection; food poisoning; scaled skin
 toxic shock syndrome; ss.

OS	Staphylococcus aureus.	Location/Qualifiers
FH	Key	

FT	key	location/qualifiers	complement (65. .511)
FT	CDS		

```

FT W09730070-A1. /*tag= a
PN 21-AUG-1997.
PD 19-FEB-1997. U02318.
PR 20-FEB-1996; US-011888.
PA (SMK ) SMITHKLINE BEECHAM CORP.
PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR WPI: 97-424969/35.
DR P-PSDB; W28082.
PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
to isolate antimicrobial compounds, and in vaccines against S.
aureus infection
Claim 9; Page 823; 989pp; English.
CC The present sequence encodes a Staphylococcus aureus protein of
CC unknown function. The present sequence was isolated from a
CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC sequence can be used in the construction of ribozymes and antisense
CC sequences to control the expression of Staphylococcal genes. The DNA
CC sequence is also useful as a source of regulatory elements for the
CC control of bacterial gene expression. The encoded protein may be used
CC to produce vaccines to enable a host to produce specific antibodies
CC with antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.
CC Sequence 584 BP; 167 A; 109 C; 84 G; 221 T;
SQ

```

alignment_scores:

Quality:	128.50	Length:	123
Ratio:	1.736	Gaps:	6
Percent Similarity:	60.163	Percent Identity:	32.520

alignment_block:

US-09-252-691-7056 x T84021/rev ..

Align seg 1/1 to reverse of: T84021 from: 1 to: 584

```

7 IleThrProGluAsnThrMetThrLysThrSerPheArgLysHisArgVa 23
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 ATAGAACCAGAAAAC.....GATAAGATAACACAGTTCGTGGAGAATTAAT 348

23 lGluArgPheSerSerArgGlnAlaThrArgArgThrProGluProGlnP 40
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 AGAATACATTGAAAAT.....

40 roThrArgValIleLeuPheAsnLysProTyrAspValLeuProGlnPhe 56
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 .....GTCATATATGTTAAATAGCCAAAGGTTATATA...TCAGCA 290

57 ThrAspGluAlaGlyArgSerThrLeuLysAspPheIleProValGln.. 72
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289 ACTGAAGATCATCTCAAAACCGTTATTGATTAACTACCTGAATACCA 240

73 .....GlyValTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyL 87
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 ACATCTAAATATTTCCTCGGTTGCTCTTGATAAGATACAGAAGGCG 190

87 euLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnPro 103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 TTTTATTGATAACGAATGATGGTGAITTTAATCATGATTAATGAGTCCA 140

104 GlyLysArgThrGlyLysIleTyrTyrValGlnValGluGluPro.. 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 AACAAACATCTTTCTAAATAATGAGTT...ATTTCAGGAATTCCTAT 93

120 .....AspAspAlaSer 123
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 CACTGAAGACGACAAATCA 74

```

```

seq_documentation_block:
ID V65214 standard; DNA; 3071 BP.
AC V65214;
DT 24-DEC-1998 (first entry)
DE DNA encoding a S. pneumoniae protein of unknown function.
KW Streptococcus pneumoniae protein; recombinant; gene expression;
KW DNA chip; virulence; antibody; infection; detection; treatment; ss.
OS Streptococcus pneumoniae.
PN W09826072-A1.
PD 18-JUN-1998.
PF 09-DEC-1997; U22578.
PR 13-DEC-1996; US-036281.
PA (ELIL ) LILLY & CO ELI.
PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI Mills BJ, Norris FH, Peery RB, Rostek PK, Rostek PR,
PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI Young Bellido ML;
DR WPI: 98-348529/30.
DR P-PSDB; W80138.
PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
for evaluating gene expression, and identification of virulence
genes
PS Claim 1; Pages 52-54; 333pp; English.
CC This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function. The invention provides DNA sequences (V65201 to V65304) from
CC the Streptococcus pneumoniae genome and corresponding protein sequences
CC (W09605 to W09728). A recombinant host containing a vector comprising of
CC the above nucleic acids can be used for the recombinant expression of
CC the protein sequences. The invention also provides a DNA chip having
CC arrayed on it at least 15 base pair fragment of any one or more of these
CC DNA sequences. The DNA chip can be used for identifying virulence genes
CC expression in S. pneumoniae and for identifying virulence genes in
CC S. pneumoniae. Antibodies that selectively bind to the above proteins or
CC peptide fragments can be used to treat S. pneumoniae infection. The
CC antibodies can also be used to detect S. pneumoniae cells.
SQ Sequence 3071 BP; 886 A; 588 C; 668 G; 929 T;

```

alignment_scores:

Quality:	127.00	Length:	278
Ratio:	0.984	Gaps:	13
Percent Similarity:	46.403	Percent Identity:	24.101

alignment_block:

US-09-252-691-7056 x V65214 ..

Align seg 1/1 to: V65214 from: 1 to: 3071

```

5 GlnLeuIleThrProGluAsnThrMetThrLysThrSerPheArgLysH 21
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
805 GAATTTATCGCAGATGACATGTCAAGGTTAAGCTTTTAAAGCA 854

21 sArgValGluArg.....
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
855 CGAGGTTTCTAAGGATTCCTGCGCAAGATTAAAGTTCCAGGCGGAGCTA 904

25 .....
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 TTCTGTCAATAATCAACCGCAAAATGCAACGTATCTATTGGACGTTGGA 954

26 .....PheSerSerArgG 30
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
955 GACTACGTTACCATTCAGATTCCTCGGTGAGAAAGCTTTGAACCTTGGGA 1004

30 nAlaThrArgArgThrProGlu.....ProGlnProThrArgValIleL 45
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1005 GCGTATTGACCTCCATTAGATATTCTCTATCAGGATGACCACTTTCTAG 1054

45 euPheAsnLysProTyrAspVal.....LeuProGlnPheThrAspGlu 59
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1055 TCTTGAATAAACCCCTATGAGTGGCTTCTATTCTTAGTGTAACTACTCT 1104

60 AlaGlyArgSerThrLeuLysAspPheIle.....

```

```

1105 .....AATACCATGCCAATTTTATCAAGGGTACTATGTCACGCA 1145
70 .....ProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAsps 84
1146 AATATTATGAAATCAGCAGGTTTCACATTGTTACCCAGACTAGATAGGACCA 1195
84 erGlUGlyLeuValLeuThrThrAsnAspGlyValLeuGlnAlaArgLeu 100
1196 CTTCTGGCTTGATGCTCTTTGCCAAGCAGCGTTATGCCCATGCACGATTA 1245
101 ThrGln.....ProGlyLysArgThrGlyLysIleTyrValGlnVa 115
1246 GACAGCAGTTGCGAAGAATATCTATCGAAGACGCTACTTTGCTTGGT 1295
115 lGlUGly.....GluProAspAspAlaSerLeuAlaLysLeu 127
1296 TAAGGAGATGGACATTGGAGCCAGAGGGGAAATATTGCTCCGATTG 1345
128 ArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleG 144
1346 CGC.....GTCATGAAGATTCCATTATTACCA..... 1372
144 uArgValAsnGluProGlnTrpLeuTrpProArgAsnProIleArgG 161
1373 .....GACGAGTGGCTTAAAGGGGAAAGTATGCCCATATTC 1409
161 luArgLysSerIleProThrSerTrp.....LeuLysIle 172
1410 ATAC.....AGATTGAGCTTCATTATGGAATATCATCTGGTCTATATT 1455
173 ThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisVa 189
1456 CACCTGCACACTGTCGAACCATCAATCCGAGTCCCATTTTCTCATAT 1505
189 lGlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrL 206
1506 CGGTTTCTTGTCTGGGA..... 1524
206 euAspSerLeuAlaAsnGlyGluTrpArgSp 216
1525 ..GATGATTGATGGTGGTAGTCTGGAAGAT 1554
seq_name: N_Geneseq_36:V52139

```

seq_documentation_block:

```

ID V52139 standard; DNA; 20199 BP.
AC V52139;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:6.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WQ9818931-A2.
PR 07-MAY-1998.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-27225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1: Page 183-194; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae

```

```

CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 20199 BP; 6009 A; 3655 C; 4458 G; 6077 T;

```

alignment_scores:

```

Quality: 126.00 Length: 278
Ratio: 0.992 Gaps: 13
Percent Similarity: 45.683 Percent Identity: 24.460

```

alignment_block:

```

US-09-252-691-7056 x V52139 ..

```

```

Align seg 1/1 to: V52139 from: 1 to: 20199

```

```

5 GlnLeuIleThrProGluAsnThrMetThrLysThrSerPheArgLysH1 21
17566 GAATTTATCGCAGATGAACATGTCAGAGTTAAGACCTTCTTAAAAAAGCA 17615
21 sarGValGluArgPheSerSerArgGlnAlaThrArgArg..... 34
17616 CGAGGTTTCTAAGGGATTGCTGGCCAAAGATTAGTTTCGAGGTGGAGCTA 17665
35 .....ThrProGluProGlnProThrArgValIleLeu..... 45
17666 TTCTGTGCAATAATCAACGCCAAATCAACGTATCTATTGGACGTTGGA 17715
45 ..... 45
17716 GACTACGTTACCATTGACATTCCCGCTGAGAAAGGCTTTGNAACCTTGA 17765
45 ..... 45
17766 GGCTATTGAGCTTCCATTAGATATCTCTATGAGGATGACCACCTTCTAG 17815
46 ..PheAsnLysProTyrAspVal.....LeuProGlnPheThrAspGlu 59
17816 TCTTGAATAAACCCATGAGTGGCTTCTTATCTCTAGTGTCAATCACTCT 17865
60 AlaGlyArgSerThrLeuLysAspPheIle..... 69
17866 .....AATACCATTGCCAAATTTATCAAGGGTACTATGTCACAGCA 17906
70 .....ProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAsps 84
17907 AATATTATGAAATCAGCAGGTTTCACATTGTTACCCAGACTAGATAGGATA 17956
84 erGlUGlyLeuValLeuThrThrAsnAspGlyValLeuGlnAlaArgLeu 100
17957 CTTCTGGCTTGATGCTCTTTGCCAAGCAGCGTTATGCCCATGCACGATTA 18006
101 ThrGln.....ProGlyLysArgThrGlyLysIleTyrValGlnVa 115
18007 GACAGCAGTTGCGAAGAATATCTATCGAAGACGCTACTTTGCTTGGT 18056
115 lGlUGly.....GluProAspAspAlaSerLeuAlaLysLeu 127
18057 TAAGGAGATGGACATTGGAGCCAGAGGGGAAATATTGCTCCGATTG 18106
128 ArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleG 144

```

```
18107 CGC.....GTGATGAAGATTCATATTACCA..... 18133
|||      |||      |||      |||      |||
144 uArgValAsnGluProGluTyrProArgAsnProProIleArgG 161
|||||||      |||      |||      |||      |||
18134 .....GACGAGTGGCTAAGCGGAAAGTATGCCCATACTTC 18170
161 luArgLysSerIleProThrSerTrp.....LeuLysIle 172
::      ::      ::      ::      ::
18171 ATAC....AAGATTGTAGCTTCTTATGGAATATTCACCTGGTCTATAT 18216
173 ThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisVa 189
|||      |||      |||      |||      |||
18217 CACCTGCACACTGGTCGACCACTCAATCCGAGTCCATTTTCTCATAT 18266
189 lGlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrL 206
:|||||||      |||
18267 CGGTTTTCCTTTGCTGGA..... 18285
206 euAspSerLeuAlaAsnGlyGluTyrArgAsp 216
|||      |||      |||      |||      |||
18286 ..GATGATTGTATGCTCTAGTCTGGAAGAT 18315
```

OM of: US-09-252-691-7056 to: Issued_Patents_NA.* out_format : pfs

Date: Jun 10, 2000 1:37 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame-p2n.model -DBV=xlp
-Q/cgn2_1/USPTO_spool/US09252691/runat_05062000_101650_615/app_query.fasta.1
-DB-Issued Patents_NA -OFMT=fastp -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.500 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=100000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:

Query: US-09-252-691-7056
Query length: 222
Database: Issued_Patents_NA.*
Database sequences: 230463
Database length: 6492525
Search time (sec): 26.040000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-700-576-1 +		206.00	391.54	2.4e-14	3061 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-247-901C-1 -		94.50	126.08	14.84	50341 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-075-904-1 -		94.50	126.08	14.84	50341 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-952-817-8 -		86.00	142.14	1.89	2639 !
/cgn2_6/ptodata/1/ina/backfiles1.seq:5210025-1 +		86.00	142.14	1.89	2639 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-804-227C-1 +		85.50	108.98	132.97	43280 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-07-998-289B-7 +		83.50	128.56	10.80	5461 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-338-702-7 +		83.50	126.54	13.99	6513 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-337-339-7 +		83.50	126.54	13.99	6513 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-724-095-7 +		83.50	126.54	13.99	6513 !
/cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:PCT-US95-14262-7 +		83.50	126.54	13.99	6513 !
/cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:PCT-US95-14378-7 +		83.50	126.54	13.99	6513 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-781-802-7 -		79.50	126.52	14.03	3147 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-804-227C-7 +		79.50	96.14	690.37	44377 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-804-198-1 +		79.50	96.14	690.37	44377 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-438-439C-1 +		79.00	137.03	3.65	1150 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-705-245-19 +		79.00	137.03	3.65	1150 !
/cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:PCT-US96-08665-1 +		79.00	137.03	3.65	1150 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-259-148A-9 +		79.00	135.52	4.42	1311 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-484-054-9 +		79.00	135.52	4.42	1311 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-876-941A-9 +		79.00	135.52	4.42	1311 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-876-941A-2 -		79.00	134.61	4.97	1419 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-194-981B-2 -		79.00	134.12	5.30	1482 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-194-981B-1 -		79.00	134.12	5.30	1482 !
/cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:PCT-US95-13703-3 +		79.00	132.90	6.19	1647 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-240-049B-1 +		79.00	130.77	8.13	1983 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-259-148A-3 +		79.00	130.40	8.53	2049 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-484-054-3 +		79.00	130.40	8.53	2049 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-876-941A-3 +		79.00	130.40	8.53	2049 !
/cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:PCT-US95-13703-1 +		79.00	130.40	8.53	2049 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-259-148A-1 +		79.00	130.15	8.81	2094 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-054-1 +		79.00	130.15	8.81	2094 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-876-941A-1 +		79.00	130.15	8.81	2094 !
/cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:PCT-US93-08849A-4 +		79.00	116.02	53.92	7168 !
/cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq:PCT-US93-08849-4 +		79.00	116.02	53.92	7168 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-808-793-24 +		77.50	113.97	70.12	6519 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-696-834-5 -		77.00	116.87	48.37	4624 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-125-468-1 -		77.00	95.40	758.86	30001 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-474-933-1 -		77.00	95.40	758.86	30001 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-651-136C-5 -		76.50	134.27	5.19	927 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-635-121-1 +		76.50	114.62	64.52	5134 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-173-508-3 -		75.50	122.34	23.99	2185 !

/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-265-310-3 - 75.50 122.34 23.99 2185
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-745-206A-6 + 75.50 110.93 103.66 5904
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-3 + 75.50 110.93 103.66 5904
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-193-078B-3 + 75.50 110.93 103.66 5904

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-700-576-1

seq_documentation_block:

; Sequence 1, Application US/08700576
; Patent No. 5750848
; GENERAL INFORMATION:
; APPLICANT: Kruger, Niels
; APPLICANT: Steinbuechel, Alexander
; TITLE OF INVENTION: DNA SEQUENCE USEFUL FOR THE PRODUCTION
; TITLE OF INVENTION: OF POLY-3-HYDROXYALKANOATES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,576
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/787-1400
; TELEFAX: 713/789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3061 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-700-576-1

alignment_scores:

Quality: 206.00 Length: 189
Ratio: 1.856 Gaps: 6
Percent Similarity: 58.730 Percent Identity: 32.275

alignment_block:

US-09-252-691-7056 x US-08-700-576-1

Align seg 1/1 to: US-08-700-576-1 from: 1 to: 3061

39 GlnProThrArgValIleLeuPheAsnLysProTyAspValLeuProG1 55
46 CCGCCGGCCGGCTACTGATGTGTACAGCCCGCTGCTGGTAAACGGC 95
55 nPheThrArgGluAlaGlyArgSerThrLeuLysAspPheIleProval. 71
96 CACCCAGCATCGCAACACCGCTACCGTCTC...GACCTGCTCCAGCGG 142
72GlnGlyValTyAlaAlaGlyArgLeuAspArgSerGlu 85
143 CGTTGCGAGTACGCTACATAGCGCGCGCTGCTCAACACCGC 192
86 GlyLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrG1 102
-193 GCGCTGATGATCTTGACCAACGATGGCCAATGGTCACGCGCGCTGACCA 242

151TriLeuTrpPro..... 154
6641 TGGCGAAGACGTGAGGCGTCGATGATGGCTCGGCCATCGCATTC 6592
155ArgAsnP 157
6591 TCCGAGTCGACGCTGCCAAGCTCGATGAGCTGGCGGATGACGGAGTC 6542
157 ro.....Pro 158
6541 CTTACCTCCGATGATCGCTAGCGCTCATCAGCATTCGGCCAATGGCCA 6492
159 lIleArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTy 175
6491 CTTGCTACTGCTGATGCTGCTCCCTTGTGAGGCGGCTAGATGTCGGCG 6442
175 rGlucGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPheP 192
6441 TCGTGGCGGTAGTCGGGACCTCAAGCGGTACTGCTCGA...GGTTCTC 6395
192 ro 192
6394 CA 6393

seq_name: /cnp2_6/ptodata/1/ina/5D_COMB.seq:us-09-075-904-1

seq_documentation_block:

Sequence 1, Application US/09075904
Patent No. 5994137
GENERAL INFORMATION:
APPLICANT: Jacobs, et al.
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebnstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,904
FILING DATE: May 11, 1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,901
FILING DATE: May 23, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Bogosian, Elizabeth A
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/475
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 50341
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: L5 shuttle phasmid sequence
DESCRIPTION: L5 shuttle phasmid sequence
HYPOTHETICAL: No
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: L5 mycobacteriophage

STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: No. 5994137e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-09-075-904-1
alignment_scores:
Quality: 94.50 Length: 184
Ratio: 1.243 Gaps: 8
Percent Similarity: 41.304 Percent Identity: 25.543
alignment_block:
US-09-252-691-7056 x US-09-075-904-1/rev ..
Align seg 1/1 to reverse of: US-09-075-904-1 from: 1 to: 50341

62 ArgSerThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaG1 78
|||||.....
6914 GCCAGACGGTTCGGATCGTCATGCCCGCCAGGGGTATTACGTACTCGG 6865
78 YargLeuAspArgAspSer.....GluGlyLeuLeu.....ValL 90
|||||.....
6864 AGACCACTTCAGGACTCCCGCAGCGGAGGGTCTTCAGGCAGATACCAC 6815
90 eutThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArg 106
|||||.....
6814 TCACCATCGACCGCTCTTCGCGAGGCGCGCGATGATGTGAGGGGAGA 6765
107 ThrGlyLysIleTyrTyrValGlnValGluGlyGluProAspAlase 123
|||||.....
6764 CGGGCAGAGCTCCGGTGGTGTGTGTGAGGCTCACCCC.....GCCTC 6721
123 rLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuP 140
|||||.....
6720 CTTTCTCAGATTCGCTAGGA.....ACTCGAC 6692
140 roAlaGlyIleGluArgValAsnGluProGlu..... 150
|||||.....
6691 CGCGGGAAGCGGTTGTACAAATCGGCCAGCTCGGCGCTTGGGCTCGACT 6642
151TriLeuTrpPro..... 154
6641 TGGCGAAGACGTGAGGCGTCGATGATGGCTCGGCCATCGCATTC 6592
155ArgAsnP 157
6591 TCCGAGTCGACGCTGCCAAGCTCGATGAGCTGGCGGATGACGGAGTC 6542
157 ro.....Pro 158

```
6541 CTTACTCCGATGATCCCTAGCGCTCATCATCAGATTCCGCGCAATGGCCA 6492
|||||
159 IleArgluArgLysSerIleProThrSerTyrLeuLysIleThrLeuTy 175
|||||
6491 CCTCGTACTCGTATGTCGCTCCCTTGTGAGGCGGTACATGTCGCCG 6442
|||||
175 rGlulGlyArgAsnArgGlnValArgMetThrAlaHisValGlyPheP 192
|||||
6441 TCGTGCCTAGTTCGCGGACCTCAAGACCGTACTGCTCGA...GGTTCTC 6395
192 ro 192
|||
6394 CA 6393
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-952-817-8

```
seq_documentation_block:
; Sequence 8, Application US/07952817
; Patent No. 5356796
; GENERAL INFORMATION:
; APPLICANT: Keller, John W.
; TITLE OF INVENTION: A Repressor Protein and Gene for Regulating
; TITLE OF INVENTION: Expression of Polypeptides and Its Use in the Preparation of
; TITLE OF INVENTION: 2,2-Dialkylglycine Decarboxylase of Pseudomonas Cepacia
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,817
; FILING DATE: 19920928
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 01120.0002-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2639 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-952-817-8
```

alignment_scores:
Quality: 86.00 Length: 221
Ratio: 0.789 Gaps: 7
Percent Similarity: 49.321 Percent Identity: 24.434

alignment_block:

US-09-252-691-7056 x US-07-952-817-8/rev ..

Align seg 1/1 to reverse of: US-07-952-817-8 from: 1 to: 2639

```
11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
|||||
1991 TCGTCAATACGACGCTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1951
```

```
27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
|||||
1950 CTCACGCCGCTCGACACGGTGGTGGCTGTTGGAACCAACACGATCGG 1901
|||||
44 leuPheAsnLysProTyrAspValLeuProGlnPhe..... 56
|||||
1900 CAGGATTGGGTGACGACGGCGGTGCTGCCGAGTTTCGCCGCGCGTGG 1851
|||||
57 .....ThrAspGluAlaGlyArgSerThrLeuLysAspPheI 69
|||||
1850 GCCGGCGTATCGTAACGTCACGTCGAGCGTCACGCTGAAG...GTGTT 1804
|||||
69 eProValGlnGlyValTyrAlaAlaGlyArg.....LeuAspArgAsp 84
|||||
1803 GCCGTGGTGGCGGATACCGGCGCAGCAAGCGGCGGTCAATGCTTCA 1754
|||||
84 exGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
|||||
1753 CCGAATCGATGCGGTGCGAATCGAATCGAATCGAATCGAATCGAATCG 1704
|||||
101 ThrGlnProGlyLys..... 105
|||||
1703 GTGCTGCCGGCGCGTGCACCGGACACGCGTTTCGGCGACACGCGCGT 1654
|||||
106 .....ArgThrGlyLysIleTyrValGlnValGluG 117
|||||
1653 CGCACATGATGCTGCTCGAACACACGAGGCTTATGCGAGTTTCGCCG 1604
|||||
117 LyGluProAspAspAlaSerLeuAlaLys..... 126
|||||
1603 GCCGTGGCGGAGGATGCTGATCGTGGGCGCGGATCACCCGTCGCGAG 1554
|||||
127 ...LeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAla 142
|||||
1553 TGTTCGCCGACGCGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1504
|||||
142 yLeuGluArgValAsnGluProGluTrpLeuTrpProArgAsnProPro 159
|||||
1503 GCATTCGCGCGGTGCGGATCGGAAAGCGTGGGCGGCGGCGGCGGCG 1454
|||||
159 LeArgGluArgLysSerIleProThrSerTyrLeuLysIleThrLeuTy 175
|||||
1453 TCGATCGGGAATCGTCCGCGGCTGCTCGAACCAGCGCGCGCGCGCG 1410
|||||
176 GluGlyArgAsnArgGlnValArgMetThrAlaHisValGlyPhePr 192
|||||
1409 .....AATCGACGATCGGAGAGGCGGTTACCGCGCGCGCGCGGT 1369
|||||
192 oThrLeuArgLeu 196
|||||
1368 TACAGCGCGCGTT 1356
```

seq_name: /cgn2_6/ptodata/1/ina/backfiles1.seq:5210025-1

seq_documentation_block:

```
; Patent No. 5210025
; APPLICANT: KELLER, JOHN W.
; TITLE OF INVENTION: REPRESSOR PROTEIN GENE FOR REGULATING
; EXPRESSION OF POLYPEPTIDES AND ITS USE IN THE PREPARATION OF
; 2,2-DIALKYLGLYCINE DECARBOXYLASE OF PSEUDOMONAS CEPACIA
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/501,814
; FILING DATE: 30-MAR-1990
; SEQ ID NO: 1
; LENGTH: 2639
5210025-1
```

alignment_scores:
Quality: 86.00 Length: 221
Ratio: 0.789 Gaps: 7

```
Percent Similarity: 49.321 Percent Identity: 24.434
alignment_block:
US-09-252-691-7056 x 5210025-1/rev ..
Align seg 1/1 to reverse of: 5210025-1 from: 1 to: 2639
11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPhe 27
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1991 TCGTCAATAACGCGGCTTCGCGCGCGCGCGCGCGCGCGCGCG 1951
27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1950 CTCAGCGCGCTCGACACGGTCGGTGGCTGTTCGAACCAACACGATCGG 1901
44 leLeuPheAsnLysProTyrAspValLeuProGlnPhe..... 56
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1900 CACGATTCGGGTGACGCGAGCGCGCTGCTCGCGCATTTTCGCGCGGTGG 1851
57 .....ThrAspGluAlaGlyArgSerThrLeuLysAspPheII 69
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1850 GCCGCGGTGATCGTGAACGTCACCTCGAGCTCAGCTGAAG...GTGTT 1804
69 eProValGlnGlyValTyrAlaAlaGlyArg.....LeuAspArgAspS 84
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1803 GCCGCTGTCGGCGCATACCGCGGCGCAGGCGCGGTCAATCGCTTCA 1754
84 erGluGlyLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1753 CCGAATCGATCGGTGCGTCAACTCGAACCGTTTCGGCGTTCGCGCGCATCTC 1704
101 ThrGlnProGlyLys..... 105
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1703 GTGTCGCGCGCGCTGTCACCGCGGACACCGCTTCGGCGACACGCGGTG 1654
106 .....ArgThrGlyLysIleTyrTyrValGlnValGluG 117
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1653 CGCATATCGATGGCTTCAACACAGCGGCTTATCGGAGTTTCGCGGAAG 1604
117 yGluProAspAspAlaSerLeuAlaLys..... 126
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1603 GCCGTGCGGAGGATCTCGATGCTGCGGCGCGATCACCGCTGCGCAGGA 1554
127 ...LeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaG 142
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1553 TGTTCGCGACCGCTGTGGCGCGCGCGGCGACGCGCTCGAGCGCGATGC 1504
142 yIleGluArgValAsnGluProGluTyrLeuTrpProArgAsnProI 159
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1503 GCATTCGCGCGCGTCCGATGCCGAGCGGTGGCGCGCGAAGCGCGTGA 1454
159 leArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTr 175
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1453 TCGATCGGGAATCCGTCGCGCGGTCGCTCGAACCGCGCGCGACG..... 1410
176 GluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhePr 192
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1409 .....NATCGCATGTCGGAAGAGCGGTACCGCGCGCGCGGT 1369
192 oThrLeuArgLeu 196
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1368 TACACGCGCGGTT 1356
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-804-227C-1
seq_documentation_block:
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
```

```
;; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
;; NUMBER OF SEQUENCES: 15
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: THOMAS G. PLANT 1501
;; STREET: LILLY CORPORATE CENTER
;; CITY: INDIANAPOLIS
;; STATE: IN
;; COUNTRY: USA
;; ZIP: 46285
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: ASCII(DOS) text only
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/804,227C
;; FILING DATE: February 21, 1997
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Plant, Thomas, G.
;; REGISTRATION NUMBER: 35,784
;; REFERENCE/DOCKET NUMBER: X-8231
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 317-276-2459
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 43280 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 816..14234
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 14351..19945
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 20010..31199
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 31232..36067
;; NAME/KEY: CDS
;; LOCATION: 36249..41774
;; US-08-804-227C-1

alignment_scores:
Quality: 85.50 Length: 202
Ratio: 0.891 Gaps: 9
Percent Similarity: 47.525 Percent Identity: 25.248

alignment_block:
US-09-252-691-7056 x US-08-804-227C-1 ..
Align seg 1/1 to: US-08-804-227C-1 from: 1 to: 43280
19 ArgLysHisArgValGluArgPheSerArgGlnAlaThrArgTh 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41806 CGGCGGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 41855
41856 GCAGGCTCCAACTACCGCGCGCGCGCGCGCGCGCGCGCGCA 41901
51 spValLeuProGlnPheThrAspGluAlaGlyArg.....Ser 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41902 ..GGGAGACCCCTACGGGATGATCTCGCGCGCGCGCGCGCG 41949
64 ThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGly 80
:..:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```



```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/338,702
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wallen III, John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19338
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6513 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-338-702-7

alignment_scores:
Quality: 83.50 Length: 133
Ratio: 1.265 Gaps: 6
Percent Similarity: 49.624 Percent Identity: 30.075

alignment_block:
US-09-252-691-7056 x US-08-338-702-7
..
Align seg 1/1 to: US-08-338-702-7 from: 1 to: 6513

11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
:::|||||::: ||| ::::: ||| :::
2022 AGCAATGACCAAGGAGAGCAAAATTCGCCAACCGCAACGCAATCA 2071
:::|||||::: ||| ::::: ||| :::

27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
:::|||||::: ||| ::::: ||| :::
2072 ATCATGGCGGCCCAATGCGGCGCACCACTGTCTGGACACCAATCA 2121
:::|||||::: ||| ::::: ||| :::

44 leuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAla 60
||| ::::: ||| ::::: ||| ::::: ||| :::
2122 AGCTGCATCATCGGCACGACGAAATTCGCTGGAGTGACGACGACGAGCT 2171
||| ::::: ||| ::::: ||| ::::: ||| :::

61 GlyArgSerThrLeuLysAsp.....PheIle...ProValGlnGlyVa 74
||| ::::: ||| ::::: ||| ::::: ||| :::
2172 GCGAAGATTAAACATCATGCAATCTTTATCGAGCCGCTCCAGACACA 2221
||| ::::: ||| ::::: ||| ::::: ||| :::

74 lTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuValLeu 91
||| ::::: ||| ::::: ||| ::::: ||| :::
2222 AACGGTGGTT.....GATATGAAGATGTGATGGTCTG. 2255
||| ::::: ||| ::::: ||| ::::: ||| :::

91 hrAsnAspGlyValLeuGlnAla.....ArgLeuThrGlnProGlyLys 105
||| ::::: ||| ::::: ||| ::::: ||| :::
2256 ..AATGACATCATCGAACAGCGCTGTGCGCACAGTCGCGCAACGCGAT 2303
||| ::::: ||| ::::: ||| ::::: ||| :::

106 ArgThrGlyLysIleTyrValGlnValGluGlyGluProAspAla 122
||| ::::: ||| ::::: ||| ::::: ||| :::
2304 CCGGGTGTCTCGTTACTATTTCACACAGAGGACGACGAG..... 2348
||| ::::: ||| ::::: ||| ::::: ||| :::

122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr 138
||| ::::: ||| ::::: ||| ::::: ||| :::
2349 .....GATGGCGCGACG 2360

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:us-08-337-339-7

seq_documentation_block:
; Sequence 7, Application US/08337339
; Patent No. 5593864
; GENERAL INFORMATION:
; APPLICANT: Warmke, Jeffrey W.
```



```

Align seg 1/1 to: PCT-US95-14262-7 from: 1 to: 6513

11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
   :::::::::::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2022 AGCAATGACCAAGAGAGCAAAATTCGCAACCGCAACACACCAATCA 2071
   :::::::::::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2072 ATCAGTGGCGCCCAACCAATGCGGCACCACTGTCTGGACCAATCACA 2121
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

44 leuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAla 60
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2122 AGCTCGATCATCGACAGCCGCTGTGCGGCACAGCGGCAAGCGAT 2171
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

74 lTyrAlaAlaGlyArgLeuAspArgSerGluGlyLeuValLeuT 91
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2222 AACGGTGGTT .....GATATGAAAGATGTGATGCTCTG. 2255
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

91 hrAsnAspGlyValLeuGlnAla .....ArgLeuThrGlnProGlyLys 105
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2256 ..AATGACATCATCGAAGAGCCGCTGTGCGGCACAGTCGCGGCAAGCGAT 2303
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

106 ArgThrGlyLysIleTyrValGlnValGluGlyGluProAspAspAl 122
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2304 CGCGGTCTCCGTTTACTATTCCCAACAGAGGACGATGACGAG..... 2348
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProthr 138
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2349 .....GATGGCGCGACG 2360
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: PCT-US95-14378-7

seq_documentation_block:
; Sequence 7, Application PC/TUS9514378
; GENERAL INFORMATION:
; APPLICANT: Warmke, Jeffrey W.
; APPLICANT: Hall, Linda
; APPLICANT: Feng, Gouping
; APPLICANT: Van Der Ploeg, Leonardus
; TITLE OF INVENTION: PROCESS FOR FUNCTIONAL EXPRESSION OF THE
; TITLE OF INVENTION: PARA SODIUM CHANNEL
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Roy D. Meredith
; STREET: P.O. Box 2000, 126 E. Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14378
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meredith, Roy D.
; REGISTRATION NUMBER: 30,777
; REFERENCE/DOCKET NUMBER: 19332 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-4678
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6513 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US95-14378-7

alignment_scores:
  Quality: 83.50      Length: 133
  Ratio: 1.265      Gaps: 6
  Percent Similarity: 49.624      Percent Identity: 30.075

alignment_block:
US-09-252-691-7056 x PCT-US95-14378-7

Align seg 1/1 to: PCT-US95-14378-7 from: 1 to: 6513

11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
   :::::::::::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2022 AGCAATGACCAAGAGAGCAAAATTCGCAACCGCAACACCAATCA 2071
   :::::::::::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2072 ATCAGTGGCGCCCAACCAATGCGGCACCACTGTCTGGACCAATCACA 2121
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

44 leuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAla 60
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2122 AGCTCGATCATCGACAGCCGCTGTGCGGCACAGTCGCGGCAAGCGAT 2171
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

61 GlyArgSerThrLeuLysAsp .....PheIle...ProValGlnGlyVa 74
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2172 GCAGAGATTAAACATCATGACATCCTTTTATCGAGCCGCTCCACAGACA 2221
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

74 lTyrAlaAlaGlyArgLeuAspArgSerGluGlyLeuValLeuT 91
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2222 AACGGTGGTT .....GATATGAAAGATGTGATGCTCTG. 2255
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

91 hrAsnAspGlyValLeuGlnAla .....ArgLeuThrGlnProGlyLys 105
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2256 ..AATGACATCATCGAAGAGCCGCTGTGCGGCACAGTCGCGGCAAGCGAT 2303
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

106 ArgThrGlyLysIleTyrValGlnValGluGlyGluProAspAspAl 122
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2304 CGCGGTCTCCGTTTACTATTCCCAACAGAGGACGATGACGAG..... 2348
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProthr 138
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2349 .....GATGGCGCGACG 2360
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq: US-08-781-802-7

seq_documentation_block:
; Sequence 7, Application US/08781802
; Patent No. 5969121
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FONSTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; --COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781.802
FILING DATE: 10-JAN-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/694,078
FILING DATE: 07-AUG-1996

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/019,580
FILING DATE: 12-JUN-1996

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/009,704
FILING DATE: 11-JAN-1996

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,995
FILING DATE: 01-AUG-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 95,963-E

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 3147 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 245..1231

OTHER INFORMATION: /note= "TSPA E101 sequence longest

open reading frame; other possible start codons are TTG/leu9;

OTHER INFORMATION: TTG/leu13; TTG/leu15; GTG/val43"

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 245..1231

US-08-781-802-7

alignment_scores:

Quality: 79.50 Length: 128
Ratio: 1.242 Gaps: 9
Percent Similarity: 50.000 Percent Identity: 31.250

alignment_block:

US-09-252-691-7056 x US-08-781-802-7/rev ..

Align seg 1/1 to reverse of: US-08-781-802-7 from: 1 to: 3147

31 AlaThrArgArgThrProGluProGlnProThrArgValIleLeuPheAs 47

1436 GCCCGCTAAGACCCGCCAACCCCGCCCTCAGC..... 1404

47 nLysProTyrAspValLeuProGlnPheThrAspGluAlaGlyArgSerT 64

1403GCCTATGACCGGCTC.....CGCACCA 1382

64 hrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGlyArgLeu 80

1381 GCCTGCGGACGCCCTCTACTTCCAGGGGTCTCGCC.....CTG 1341

81 AspArgAspSerGluGlyLeuValLeuValLeuThrAsnAspGlyValLeu 97

1340 AAGCGGAC.....GTCTGTGGCGAAGGGGAGGGGTCTCTCGG 1300

97 nAlaArgLeuThrGlnProGlyLysArgThrGlyLysIleTyrValG 114

..... 114

1299 GGCCCTCTCTCAGCGCCCTCTCCCAAG.....GAGCGGTACGGGTCC 1256

114 InValGluGlyGluProAspAlaSerLeuAlaLysLeuArgAsnGly 130

1255 TCCTCGAGGCTCTCCCGAC.....CTCAGGACCTACGGGGGGG 1215

131 ValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGluArgValas 147

1214CCTGCCCTCTCTCCCGACGGCTCTCGAGCC..... 1185

147 nGluProGluTyrLeuTyrProArgAsnProPro 158

1184 .AGTCCAGAA.....CGCTACCTCTCC 1164

seq_name: /cgn2_5/ptodata/1/ina/5C_COMB.seq:US-08-804-227C-7

seq_documentation_block:

Sequence 7: Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: DeHoff, Bradley S.

APPLICANT: Kustoss, Stuart A.

APPLICANT: Rostock, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 44377 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 350..14002

FEATURE:

NAME/KEY: CDS

LOCATION: 14046..20036

FEATURE:

NAME/KEY: CDS

LOCATION: 20110..31284

FEATURE:

NAME/KEY: CDS

LOCATION: 31329..36071

FEATURE:

NAME/KEY: CDS

LOCATION: 36155..41830

US-08-804-227C-7


```
alignment_scores:
  Quality: 79.50      Length: 215
  Ratio: 0.914       Gaps: 10
  Percent Similarity: 40.465  Percent Identity: 24.186

alignment_block:
US-09-252-691-7056 x US-08-804-227C-7

Align seg 1/1 to: US-08-804-227C-7 from: 1 to: 44377

3 MetArgGlnLeuIleThrProGluAsnThrMethThrLysThrSerPheAr 19
  ::::::::::::::::::::
33596 CTTCCGGGACACTCACCCAGGTGGAGCA..... 33625

19 gLysHisArgValGluArgPheSerArgGlnAlaThrArgThr 36
  ::::::::::::::::::::
33626 .....GTTCCGGCCGAGCTCTCCGACAGCTGGGCGCGTCGCG 33665

36 roGluProGlnProThrArgValIleLeuPheAsnLysProTyrAspVal 52
  ::::::::::::::::::::
33666 CCGACCGCGCGCGG..... 33679

53 LeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAspPheI1 69
  ::::::::::::::::::::
33679 ..... 33679

69 eProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgSerGluG 86
  ::::::::::::::::::::
33680 .....GCTGCCCTTCTCTCCACGCTGACCGCGCG 33710

86 lyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln 102
  ::::::::::::::::::::
33711 GC.....GCACGAGCCCGAAG 33727

103 ProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGly..... 117
  ::::::::::::::::::::
33728 CTTGGACCGCCGCTACTGCTACCGGACATCGCGGACCCCGTGAGTCTG 33777

118 ....GluProAspAlaSerLeu.....AlaLysLeuArgAsnG 130
  ::::::::::::::::::::
33778 CGTCCACCTCGGACGCTGCTCGGAGGCGCCACCGACCTTCGT.... 33823

130 lyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGlu..... 144
  ::::::::::::::::::::
33824 .....CGAGATGGCGCCACACCCCTGCTGGCGCGCGATCGA 33862

145 .....ArgValAsnGluProGluTyrLeuTyrPpr 154
  ::::::::::::::::::::
33863 CGAGGTCCGCGAGCGCGAGGCGGTGACGCGCCCGC.....CC 33900

154 oArgAsnProProlleArgGluArgLysSerIleProThrSertTrpLeuL 171
  ::::::::::::::::::::
33901 TCGCACCTCCACCGCGCTCGCGGCGCTGGACCGGTTCGCTCTC. 33949

171 ysIleThrLeuTyrGluGlyArgAsnArgGlnValArgArgMethThrAla 187
  ::::::::::::::::::::
33950 .....GGTGGGCGCGCTTCGCGCCACGCGACGACGAGTACGGGT 33982

188 HisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly 202
  ::::::::::::::::::::
33983 CGACTGGACCGCCCT...CTTCGAGGCTCCGCGCGCGCGCGGGT 34024

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq.us-08-804-198-1
seq_documentation_block:
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kubstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.

; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350...14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046...20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110...31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329...36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155...41830
; US-08-804-198-1

alignment_scores:
  Quality: 79.50      Length: 215
  Ratio: 0.914       Gaps: 10
  Percent Similarity: 40.465  Percent Identity: 24.186

alignment_block:
US-09-252-691-7056 x US-08-804-198-1

Align seg 1/1 to: US-08-804-198-1 from: 1 to: 44377

3 MetArgGlnLeuIleThrProGluAsnThrMethThrLysThrSerPheAr 19
  ::::::::::::::::::::
33596 CTTCCGGGACACTCACCCAGGTGGAGCA..... 33625

19 gLysHisArgValGluArgPheSerArgGlnAlaThrArgThr 36
  ::::::::::::::::::::
33626 .....GTTCCGGCCGAGCTCTCCGACAGCTGGGCGCGTCGCG 33665

36 roGluProGlnProThrArgValIleLeuPheAsnLysProTyrAspVal 52
  ::::::::::::::::::::
33666 CCGACCGCGCGCGG..... 33679

53 LeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAspPheI1 69
  ::::::::::::::::::::
33679 ..... 33679

69 eProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgSerGluG 86
  ::::::::::::::::::::
33680 .....GCTGCCCTTCTCTCCACGCTGACCGCGCG 33710

86 lyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln 102
  ::::::::::::::::::::
33711 GC.....GCACGAGCCCGAAG 33727

103 ProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGly..... 117
  ::::::::::::::::::::
33728 CTTGGACCGCCGCTACTGCTACCGGACATCGCGGACCCCGTGAGTCTG 33777

118 ....GluProAspAlaSerLeu.....AlaLysLeuArgAsnG 130
  ::::::::::::::::::::
33778 CGTCCACCTCGGACGCTGCTCGGAGGCGCCACCGACCTTCGT.... 33823

130 lyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGlu..... 144
  ::::::::::::::::::::
33824 .....CGAGATGGCGCCACACCCCTGCTGGCGCGCGATCGA 33862

145 .....ArgValAsnGluProGluTyrLeuTyrPpr 154
  ::::::::::::::::::::
33863 CGAGGTCCGCGAGCGCGAGGCGGTGACGCGCCCGC.....CC 33900

154 oArgAsnProProlleArgGluArgLysSerIleProThrSertTrpLeuL 171
  ::::::::::::::::::::
33901 TCGCACCTCCACCGCGCTCGCGGCGCTGGACCGGTTCGCTCTC. 33949

171 ysIleThrLeuTyrGluGlyArgAsnArgGlnValArgArgMethThrAla 187
  ::::::::::::::::::::
33950 .....GGTGGGCGCGCTTCGCGCCACGCGACGACGAGTACGGGT 33982

188 HisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly 202
  ::::::::::::::::::::
33983 CGACTGGACCGCCCT...CTTCGAGGCTCCGCGCGCGCGGGT 34024

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq.us-08-804-198-1
seq_documentation_block:
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kubstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
```

33679 33679
69 eProValGInGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGlu 86
||||| ||| ||||| :||
33680GCTGCCCTTCTTCCACCGTGACCGCGGG 33710
86 lyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln 102
|| ||||| :||
33711 GC.....GCACGAGCCCGAAGG 33727
103 ProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGly..... 117
|||||:||||| :||
33728 CCTGGACGCGGTACTGTACCGGACATGGCGGACCCGTGGAGTTGG 33777
118GluProAspAspAlaSerLeu.....AlaLysLeuArgAsnG 130
:||||| |||||:|
33778 CGTCCACCTGCGGACCGTCTCGCGAGGGCCACGCGACCTTCGT.... 33823
130 lyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGlu..... 144
:|||||:||||| :||
33824CGAGATGGCGCGCACCCCTGCTGGCGCGCGCATCGA 33862
145ArgValAsnGluProGluTyrLeuTyrPr 154
|||:|||||:|
33863 CGAGGTCCGCGGAGCGCGGTGCACGCCACCGC.....CC 33900
154 oArgAsnProProIleArgGluArgLysSerIleProThrSerTyrLeuL 171
|||||:||||| ||| ||||| :||
33901 TCGCCACCTCCACCGCGCTCCGCGGCTGGACGGTTCGCTCCTC. 33949
171 ysIleThrLeuTyrGluGlyArgAsnArgGlnValArgMetThrAla 187
||| :||||| ||| ||||| :||
33950GCTGGCGCGCGGTTCGCCACCGAGTACGGGT 33982
188 HisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly 202
:||||| ||| ||||| :|||||
33983 CGACTGGGAGCGCCT...CTTCGAGGGCTCCGCGCGCGCGCGGT 34024


```

5  CACCAGGTAAAGCGATTACGCTCCACGCTTCTACCAAGCGCTAAACCTGA 54
37  uProGlnProThrArgValIleLeuPheAsnLysProTyrAspValLeuP 54
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
55  AAACACGCCACCGCGTGTGTCTGTCTCAATAAACCCCTAGCATGTTCTC 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
54  roGlnPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIlePro 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
105  CGCAGTTCACCGATGAAGCGGACGACGAACATTAAGAATTCATCCCG 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
71  ValGlnGlyValTyrAlaLeuAlaGlyArgLeuAspArgSerGluGlyLe 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
155  GTTCAGGGTGTATTCAGCAGCAGGTCGCTTGACCGCATAGCGAAGGTT 204
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
87  uLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProG 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
205  GCTGGTGCTGACCAATAACGCGCGGTTCGAGGCGCGTTTAAACCGACGG 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
104  lYlYsArgThrGlyLysIleTyrTyrValGlnValGluGly 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
255  GTAAACGCCACCGGAAAATCTATATCTGTCGAGGTGGAAGGT 295

```

seq_name: gb_est26:AI373988

seq_documentation_block: 283 bp mRNA EST 15-JAN-1999
 LOCUS AI373988
 DEFINITION SWOVAFCAP32B02SK Onchocerca volvulus adult female cDNA
 (SW98MLW-OVAF) Onchocerca volvulus cDNA clone SWOVAFCAP32B02 5',
 mRNA sequence.

ACCESSION AI373988
 VERSION AI373988.1 GI:4160024
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
 AUTHORS Lizotte-Waniewski, M. and Williams, S.A.
 TITLE Genes expressed in adult female stage of Onchocerca volvulus
 JOURNAL Unpublished (1998)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1797276.
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu

Seq primer: pBluescript SK.
 Location/Qualifiers
 1..283
 /organism="Onchocerca volvulus"
 /db_xref="taxon:6282"
 /clone="SWOVAFCAP32B02"
 /clone_lib="Onchocerca volvulus adult female cDNA
 (SW98MLW-OVAF)"
 /sex="female"
 /dev_stage="adult"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
 Xho I; Filarial nematode parasite of humans. Two adult
 female worms of Onchocerca volvulus were isolated from
 consenting patients and quick frozen. Adult female mRNA
 was converted to double-stranded cDNA using reverse
 transcriptase and oligo(dT) followed by RNase H and DNA
 pol I. The library has 7 x 10⁵ independent recombinants
 and the average insert size is ~1100bp. The library was
 constructed by Michelle Lizotte-Waniewski with worms
 provided by Dr. Sara Lustigman. The library is available
 from Dr. Steven A. Williams, email: genome@smith.edu."

BASE COUNT
 ORIGIN

alignment_scores:

Quality: 392.00 Length: 80
 Ratio: 5.227 Gaps: 0
 Percent Similarity: 93.750 Percent Identity: 91.250

alignment_block:

US-09-252-691-7056 x AI373988 ..

Align seg 1/1 to: AI373988 from: 1 to: 283

```

123 SerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLe 139
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3  ACACCTGAAGCCCTTGCACATGCGCTAACCTTAATGATGGCCCTACCT 52
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
139 uProAlaGlyIleGluArgValAsnGluProGluTrpLeuTrpProArgA 156
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
53 GCCCGCGCGCGGAGCTGTTGACGAACCGCGGTGTTATGGCCCGGA 102
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
156 snProPheLeuArgGluArgLysSerIleProThrSerTrpLeuLysIle 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
103 ATCCACCAATTCGTGAACGCAAAAGTATCCACCAAGCTGGCTGAAGATC 152
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
173 ThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisVa 189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
153 ACCTTATATGAAGGAGCTAATCGCCAGGTGCGCGCATGACCGCCCATGT 202
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
189 lGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly 202
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
203 TGGCTTCCCACGCTCGCACTGATTCGCTATCGATGGGT 242
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

seq_name: gb_est28:AI488035

seq_documentation_block:

LOCUS AI488035 510 bp mRNA EST 29-JUN-1999
 DEFINITION EST248357 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
 CLED19M8, mRNA sequence.

ACCESSION AI488035
 VERSION AI488035.1 GI:4383406
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 510)
 AUTHORS Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,
 Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B.,
 Fujii, C.Y., Bowman, C.E., Nierman, W., Fraser, C.M., Venter, J.C.,
 Martin, G.B., Tanksley, S.D. and Giovannoni, J.
 Generation of ESTs from tomato carpel tissue
 Unpublished (1999)

On May 18, 1998 this sequence version replaced gi:3137718.
 Contact: David Frisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@CLEMSON.EDU.

TITLE

JOURNAL

COMMENT

FEATURES

source

```

1..510
/organism="Lycopersicon esculentum"
/cultivar="TM496"
/db_xref="taxon:4081"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:"

```

XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 151 a 111 c 113 g 135 t
ORIGIN

alignment_scores:
Quality: 129.50 Length: 137
Ratio: 1.660 Gaps: 4
Percent Similarity: 56.934 Percent Identity: 26.277

alignment_block:
US-09-252-691-7056 x AI488035 ..
Align seg 1/1 to: AI488035 from: 1 to: 510

34 ArgThrProGluProGlnProThrArgValLeu...PheAsnLysPr 49
91 CGCCTTCTAGAAATTCCTACAAAGTCTATCTGCACCTAAACAAGCC 140

49 OTYrAspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuL 66
141 AAAAGGTACATATGCTCATCTCGGAGAAAGAACTAAGTCATGT 190

66 ys.....AspPhele.....
191 CCCTTTTGTGATGCTTTATATAAGAGTTGGGATAAAGGCATCTGGACAA 240

70 ProValGlnGlyValTyAlaAlaGlyArgLeuAspArgSerGluL 86
241 CCGAAACCTCGCCTCTTACAGTTGGCAGACTTGATTTGCCACGACTGG 290

86 yLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnP 103
291 CTTGATATTGTGACCATGATGAGGGAGTTCACCACAGATTTCACATC 340

103 roGlyLysArgThrGlyLysIleTyTyValGlnValGluGlyGluPro 119
341 CTTTCATCTAAATTGTCAAAGGAATACATGCAACTATCGACGGTGAAGTT 390

120 AspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspG 136
391 CATAAGCGACACTGTGATAGCCATTAGTGAGGGAACAATATTATGACGTGT 440

136 yProThrLeuProAlaGlyIleGluArgVal...AsnGluProGluTPl 152
441 CCATTGCCACCCAGATAATGTTGAACTACTACCGGGCGACCTGACTTAT 490

152 eutrPProArg 155
491 CAAGACCTCGT 501

seq_name: gb_est37:AI993553

seq_documentation_block:
LOCUS AI993553 452 bp mRNA EST 08-SEP-1999
DEFINITION 701496622 A. thaliana, Ohio State clone set Arabidopsis thaliana
CDNA clone 701496622, mRNA sequence.

ACCESSION AI993553
VERSION AI993553.1 GI:5840450
KEYWORDS EST.
SOURCE thale cress.

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE 1 (bases 1 to 452)
Chen, J., Momiya, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hillman, J., Guebler, K., Kim, C., Doyle, M., Brzoska, P.,
Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D.,

Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C.,
Carpio, I., Policky, J., Suzuki, G., Argentine, C., Shah, S.,
Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and
Hanson, D.

Arabidopsis thaliana Gene Expression MicroArray

Unpublished (1999)

On May 1, 1997 this sequence version replaced gi:2059644.

Contact: David Smoller, Ph.D.

Genome Systems, Inc., a wholly owned subsidiary of Incyte

Pharmaceuticals, Inc.

4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733

Fax: 314-427-3324

Email: service@genomesystems.com.

FEATURES

Location/Qualifiers

source

1..452

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/clone="701496622"

/note="A. thaliana, Ohio State clone set"

/note="cdna library was made from selected clones from the

Arabidopsis thaliana Ohio State clone set."

BASE COUNT 132 a 86 c 93 g 141 t

ORIGIN

alignment_scores:
Quality: 95.00 Length: 111
Ratio: 1.638 Gaps: 2
Percent Similarity: 52.252 Percent Identity: 26.126

alignment_block:

US-09-252-691-7056 x AI993553 ..

Align seg 1/1 to: AI993553 from: 1 to: 452

34 ArgThrProGluProGlnProThrArgVal...IleLeuPheAsnLysPr 49
108 CGTATTCCTAAGAACTTCCTCCAAAGGTTTATTTCTGCTGCAACAAGCC 157

49 OTYrAspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuL 66
158 TAAAGGATATTTGTTCTTCGCGAGAGAAAGAGATCAATCTGCTATTA 207

66 ysAspPhe.....
208 GTTTGTTGACGAGTACTTGTCTTGGGATAAAGGAATCCGGGACT 257

70 ProValGlnGlyValTyAlaAlaGlyArgLeuAspArgSerGluL 86
258 CCCAAACCTCGCCTTTTACTGTTGTCGCTTGATGTTGCCACAACTTCACATC 307

86 yLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnP 103
308 ATTGATAGTTGTTACAAATGATGGAGATTTCGCACAGAACTTTCACATC 357

103 roGlyLysArgThrGlyLysIleTyTyValGlnValGluGlyGluPro 119
358 CTTTCATCTAGTTTACCGAAAGATATATATACACGGTTGTCGGTGATATA 407

120 AspAspAlaSerLeuAlaLysLeuArgAsnGly 130

408 CACAACGACACTTAATGGCTATCATGTAAGGC 440

seq_name: gb_est18:AA701638

seq_documentation_block:

LOCUS AA701638 844 bp mRNA EST 19-DEC-1997

DEFINITION z143e01.s1 Soares_fetal_liver_spleen_inFLS_S1 Homo sapiens cDNA
clone IMAGE:433560 3' similar to TR:Q93639 Q93639 F30A10.5, mRNA

sequence.

ACCESSION AA701638

VERSION AA701638.1 GI:2704803

Email: est@watson.wustl.edu

seq_name: qb_est27:AI415775

seq_documentation_block:

LOCUS	AI415775	408 bp	mRNA	EST	0

AACTGGAAGAAATTAATTAAAGATCTTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

Seq primer: T3 ET from Amersham
High quality sequence stop: 395.

FEATURES

1. *Staphylococcus aureus*

```
/ub_xref=taxon:7955"
/clone lib="zebrafish WashU motif pcm"
```

```

/crown_tis=separation was no MFING EST
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield

```

stage embryos"

```
/lab_host="XLI-blue MRF"
/note="Vector: nspopt1:
```

strand cDNA was primed with a Not I - oligo(dT)15 primer [5']pGACTAGTCTAGATCGGAGCGCCCTTTTTTTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL) digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed

[5' pGACTAGTCTAGATCGCGAGCGGCCCTTTTTTTTTT3'];

double-stranded cDNA was ligated to Sal I adaptors (BRL),

digested with Not I and cloned into the Not I and Sal I

sites of the psfprf1 vector (BRL): Library was constructed

Page 6

Seq primer: -40UP from Gibco
High quality sequence stop: 466.

FEATURES

source

1. .494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:272416"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 83 a 114 c 159 g 137 t 1 others
ORIGIN

alignment_scores:
Quality: 83.00 Length: 107
Ratio: 1.596 Gaps: 5
Percent Similarity: 48.598 Percent Identity: 28.972

alignment_block:

US-09-252-691-7056 x AI681127 ..

Align seg 1/1 to: AI681127 from: 1 to: 494

135 AspGlyProThrLeuPro.....AlaGlyI 143
|||||.....
122 GACGGCCACCCCTCCAGGCGACGGGACTTGGTTTCTGCTCTCT 171
|||||.....
143 eGluArgValAsnGluProGluTrpLeuTrpPro..... 154
|||||.....
172 AGAGGTACATGGGAACAGCGTGGCTCTCCAGACTGTGTTAGCTGTG 221
|||||.....
155ArgAsnProProlIeArgGluArgLysSerIlePro..... 166
|||||.....
222 ACCGTCAGGCGAGCCCTGGGGTGGCCACTGTGTCACCAAGTGGAGG 271
|||||.....
167 ThrSerTrpLeuLysIleThrLeuTyArgGluArgAsnArgGlnVal 183
|||||.....
272 TCAGCATGGCTGNGCCACGCTGTGTGAGAGGGTCTCGCAGTGGGTACG 321
|||||.....
183 gArgMetThrAlaHsValGlyPheProThrLeuArgLeuIleArgTy 200
|||||.....
322 CAGGATGTATGGCAGGGCTCTTTTCCAGTGCTT.....G 356
|||||.....
200 laMetGlySerTyThrLeuAspSerLeuAlaAsnGlyGlu..... 213
|||||.....
357 CTCATGGCCAGCTCTCGCAGGAGCCCTGCTCTGGGGAGAGCGAGGGT 406
|||||.....
214TrpArgAsp 216
||| |||
407 GTGCCAGGCTTTGGTCCGAT 427

seq_name: gb_gss6:AQ871305

seq_documentation_block:

LOCUS AQ871305 801 bp DNA GSS 03-NOV-1999
DEFINITION nb0042J08f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
clone nb0042J08f, genomic survey sequence.
ACCESSION AQ871305
VERSION AQ871305.1 GI:6221756
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 801)
Wing,R.A. and Dean,R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
On Feb 19, 1999 this sequence version replaced gi:4130187.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 17
High quality sequence stop: 384.

FEATURES

source

1. .801
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nb0042J08f"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 167 a 248 c 236 g 148 t 2 others
ORIGIN

alignment_scores:

Quality: 82.50 Length: 239
Ratio: 0.868 Gaps: 12
Percent Similarity: 39.749 Percent Identity: 22.176

alignment_block:

US-09-252-691-7056 x AQ871305/rev ..

Align seg 1/1 to reverse of: AQ871305 from: 1 to: 801

29 ArgGlnAlaThrArgArgThrProGluProGlnPro..... 40
|||||.....
729 CGCAGCTCGACTCCCGGACAGGTGAGCCCTACTCATGCATGCAGATT 680
|||||.....
41ThrArgValIleLeuPheAsnLysPro. 49
|||||.....
679 CTCCTCTTCGGGTCTCCGTTGAACGAGATTTTGTCTTTTCAGTCCGC 630
|||||.....
50TyrAspVal 52
|||||.....

```

629 TTCGATTAGTGTGTTATGCCATTGCCCGCGGGCGGTACAGTGT 580
53 LeuProGlnPheThrAspGluAlaGly..... 61
   :::::|||||
579 GCGTGCAGTCCGATGAGGAAAGCGCGCGAATCGACTGGACAACCT 530
62 .....ArgSerThrL 65
529 GGGGTTCGGGCTGACGCCGACCGACTACATGTACGTCTGCTCGC 480
   |||||
65 euLysAspPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAsp 81
   |||||
479 TGAGAGAC.....GCGTCTTCTCCGCGCGAGCTCAGC 445
82 Arg.....AspSerGluGlyLeuLeu..... 88
444 CGCTAGCGCAACATCGAGCTCAGCCCTCTCCGGCGTCATCAACTACGG 395
89 .....ValLeuThrAsnAspGlyValLeuGlnAlaArgLeuT 101
394 CCAGGTCTGTCACTGTTCCAGTCCACACATGTCCATGCAACGAAAGC 345
101 hrGlnProGlyLysArgThrGlyLys..... 109
344 TCGATCCAGCGCAGCGGACTACAAAGTGAACGAGGCTGATGCTCGCTGC 295
110 .....IleTyrTyrValGlnValGluGlyGluProAspAlaSerLe 124
294 GTACGTACGTATGTGCTC.....GATGATCAGGGGCT 263
124 u.AlalysLeu.....ArgAsnGlyValThrLeuAsnAspGlyProThr 138
262 CTTGAGGCTCTGAAGCGCTACAGCGCGCGCAACCAACAGGGGTCGTACA 213
139 LeuProAlaGlyIleGluArgVal.....AsnGluProGln 150
212 TCGTTCGCGCGGAGGAGAACCGCGCGCGATGCAGCACCGCGCGGAG 163
150 utrPLeu.....TrpProArgAsnProPheIleArgG 161
162 CGCATGTGCATCGCTGCGCGTGGTGGAGCAGTTCGTCCACCGCGTCAA 113
161 luArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrGluGly 177
112 GCAGACCGCTCTCGCAACCGCGCTGGGTAGCGTACGTAAGCGGTACGAA 63
178 ArgAsnArgGlnVal 182
62 CGTAAACCGACCGCTC 48

```

seq_name: gb_est46:AV398064

```

seq_documentation_block: 758 bp mRNA EST 05-FEB-2000
LOCUS AV398064 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
DEFINITION AV398064 Bombyx mori cdna clone NV021042 T3, mRNA sequence.
Bombyx mori

```

ACCESSION AV398064

VERSION AV398064.1 GI:6901716

KEYWORDS EST.

SOURCE domestic silkworm.

ORGANISM Bombyx mori

```

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
Ditrysia; Bombycoidea; Bombycidae; Bombyx.

```

REFERENCE 1 (bases 1 to 758)

Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.

Bombyx mori cdna

TITLE Unpublished (2000)

JOURNAL On Jan 6, 2000 this sequence version replaced gi.6675713.

COMMENT Contact: Mita K

Genome Research Group

National Institute of Radiological Sciences

Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

```

Email: kmitta@nirs.go.jp
Method: uni-directional, sequence direction: sequenced from T3 primer
(5' -> 3')
Project: 'Silkworm Genome Program in MAFF, and Research for the
Future Program in JSPS'. see 'SilkBase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.

```

FEATURES

```

source
1..758
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021042"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/cell_type="BmN cultured cell"
/dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT 179 a 218 c 249 g 112 t
ORIGIN

```

alignment_scores:

```

Quality: 82.00 Length: 254
Ratio: 0.732 Gaps: 15
Percent Similarity: 44.094 Percent Identity: 23.228

```

alignment_block:

US-09-252-691-7056 x AV398064 ..

Align seg 1/1 to: AV398064 from: 1 to: 758

```

24 GluArgPheSerSerArg.....GlnAlaThrArgArgThrPr 36
   |||||
30 GAGGCGGGTGCAGCGCGCTGTGCGTCCGCCCGCTCAACACCGCAC 79
36 oGluProGlnProThrArgValIleLeuPheAsnLysProTyrAspValL 53
   |||||
80 CCGCGCAGAGCGCGCTCGAGCC..... 101
53 euProGlnPheThrAsp.....GluAla 60
   |||||
102 ..CGCAGCGCAGTCCGAAACCGAGGAGTGAACGCGCGGAGGCG 149
61 GlyArgSerThrLeuLysAspPhe.....IleProValGlnG 73
   |||||
150 GGGAGGCGCGCTGCGCTGTATATTCACGGGAACATGCGCGCCAAGAT 199
73 yValTyrAlaAlaGlyArgLeuAspArgAsp...SerGluGlyLeuLeu 89
   |||||
200 CACTTGGAGGAAGAGAAATCACTATCGACGCGCAGGCGCGGAGCG 249
89 alLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLys 105
   |||||
250 GCGTGTCTGCGCGCAGCGCTGGAG.....ATCGTGTCTGTACCGG 293
106 ArgThrGlyLysIleTyrTyrValGlnValGluGlyGluProAspAl 122
   |||||
294 AACGACACCGCGCTACATCTGCATCGCTGAGAACGAG.....TT 334
122 aserLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrL 139
   |||||
335 CGGCATCAGCAGGAGGAGATCCACCTGCAAGTGAACGATCCGCTGAGGA 384
139 euProAlaGlyIleGlu...ArgValAsnGlu..... 148
   |||||
395 CGCGGTGCGCATCGCGGAGAACAAACGAAGTAATAACGGGGGAGATG 434
149 .....ProGln 150
435 GGCACACCGCTCGTCTTTCGCTGCTCGTGTACGGATACCGGACCCCGGA 484
150 utrLeuTrpProArg.....AsnProIleArgGluArgLysSerI 165
   |||||
485 ATCTTCTGGTACAGAGGACCTTAACGGTCCCTATG.....G 519

```

```

Align seg 1/1 to: AA816992 from: 1 to: 805

46 PheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGlyAr 62
||||| ||||| ||||| ||||| |||||
57 TTCAATAAA...TATCAA.....ACGACGACGAGTGC CGC 88
||||| ||||| ||||| ||||| |||||
62 gSerThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGlyA 79
||||| ||||| ||||| ||||| |||||
89 CAATGAATAATCTGATCTGTCACATTTTCTTGGCGCGCTCTACGTGCC 138
||||| ||||| ||||| ||||| |||||
79 rgLeuAspArgAsp.....SerGluGlyLeuLeuValLeuThr 91
||||| ||||| ||||| ||||| |||||
139 GCTTCGCGCGAGCGGAGGTCAAGATTGAGAGGGTGTCTGTGTGCGCAGC 188
||||| ||||| ||||| ||||| |||||
92 AsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGl 108
||||| ||||| ||||| ||||| |||||
189 GTGGACAACTTCAACGACGCTGATGCCCAACAGAG..... 224
||||| ||||| ||||| ||||| |||||
108 yLysIleTyrTyrValGlnValGlu.GlyGluProAspAspAlaSerLeu 124
||||| ||||| ||||| ||||| |||||
225 .....TTCTGTCGTGGTTCGAGTCTTACGCGCCATGGTGCGACACT 264
||||| ||||| ||||| ||||| |||||

```

```

125 AlalysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAl 141
|||||TTC|||||:::|||||
265 GCAAGGTC.....TGGCTCCCGAGTACGCCAAGGCTGCCAC 302
|||||TTC|||||:::|||||
141 aGlyIleGluArgValAsnGluProGluIuThrLeuTrpProArgAsn...P 157
::: ||| :::::|||||::: ||| |||||::: |||
303 CAACTGGCCGAGAGGAGTGGCCCATCA...ACCTGGCCCAAGGTCGATGC 349
|||||TTC|||||:::|||||
157 roProIleArgGluArgGlySerIleProThrSertTrpLeuIleThr 173
|||||TTC|||||:::|||||
350 CACCGTCGAGG.....GTGAGCTGGCCGAGCAGTACG 381
|||||TTC|||||:::|||||
174 LeuTrpGluGlyArgAsnArgGluValArgArgMetThrAlaHisValGl 190
|||||TTC|||||:::|||||
382 CC.....GTCCGT.....GG 391
|||||TTC|||||:::|||||
190 yPheProThrLeuArgLeuIleArgTrpAla.....MetG 202
|||||TTC|||||:::|||||
392 CTACCCCACTCTGAAGTTCTCCGACGGGCTCTCCAGTGGAGTACAGCG 441
|||||TTC|||||:::|||||
202 lySertTrpLeuAsnSerLeuAlaAsnGlyGluTrpArgAsp 216
|||||TTC|||||:::|||||
442 GTGGCCGCACGCTGGATCATCTGCTGGGTGACCAAGAAGAC 485
|||||TTC|||||:::|||||
req_name: gb_gss15:AQ570573
req documentation block:

```

LOCUS	AQ570573	512 bp	DNA	GSS	01-JUN-1999							
DEFINITION	HS-5362_B1_G11_TTA_PCT-1.1 Human Male BAC Library Homo sapiens genomic clone FTA-938 Col-21 Row-N, genomic survey sequence.											
ACCESSION	AQ570573											
VERSION	AQ570573.1	GI:4963793										
KEYWORDS	GSS.											
SOURCE	human.											
ORGANISM	Homo sapiens											
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.											
AUTHORS	1 (bases 1 to 512)											
	Mahairs,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood L.											

JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	On Mar 23, 1999 this sequence version replaced gi:3325394. Contact: Mahairas GG, Wallace JC, Hood L

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA

ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE	1 (bases 1 to 514)
AUTHORS	Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrgg,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C. ↑
TITLE	Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL	Plant Physiol. 106, 1241-1255 (1994)
MEDLINE	95148729
COMMENT	On Apr 14, 1993 this sequence version replaced gi:838028. Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan State University MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI Tel: 517-353-0854 Fax: 517-353-9168 Email: 22313tcn@bm.cl.msu.edu

```

source
l: 374
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="170K5T7"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."
BASE COUNT      137 a      101 c      104 g      152 t      20 others
ORIGIN

alignment_scores:
      Quality:      81.50      Length:      95
      Ratio:        1.567      Gaps:      4
      Percent Similarity: 54.737      Percent Identity: 29.474

alignment_block:
US-09-252-691-7056 x R65390      ..

Align seg 1/1 to: R65390 from: 1 to: 514

34  ArgThrProGluProGlnPheThrArgVal...IleLeuPheAsnLysPr 49
|||||
108  CGTATTCTTAGAAACTCCGCCAAGGTTTATTGTGCTCGAACAGCC 157
|||||

49  oTyrAspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuL 66
|||
158  TAAAGGATATATTGTCTTCGCGAGAGAAAAGAGATCAAAATCTGCTATTA 207
|||
66  ysasppe.....Ile 69
::|||
208  GTTTGTTTGACAGTACTGTGCTCTTTGGGATAAAAGGAATCCGGGGACT 257
|||||
70  ProValGlnGlyValTyrAlaAalaglyValLeuAspArgAspSerGluGl 86
|||||
258  CCCAAACCTCGCCTTTTACTGTGTGGTCGCTCTTGATGTGCCACACGGG 307
|||||
86  yLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln. 102
|||||
308  ATTGATAGTCTTACAAATGATGAGATTTCGCACAGAAACTTTTCCATC 357
|||||

```


gb_om:BOVRP3A	7.00	101.09	2135	D13613 Bovine mRNA for rabphl	+	7.00	98.50	3139	gb_pr1:HSAA005898
gb_pat:YEC07809	7.00	101.09	2135	E07809 cDNA encoding rabphl	-	7.00	98.50	3140	em_in1:DMC102
gb_pr1:YEC07820	7.00	100.91	2191	L27698 yeast chaperonin protei	+	7.00	98.49	3145	gb_pr1:CVAC102
gb_om:SSU33211	7.00	100.90	2191	X33211 Synchococcus sp. FGAM	+	7.00	98.98	3212	gb_pr1:HSDDG1
gb_ov:S76880	7.00	100.88	2202	S76880 Xlhx-bHLH function dom	+	7.00	98.23	3266	gb_in1:DMRTGENE
gb_pr1:CHHS970	7.00	100.88	2203	X81860 C.herbarum mRNA for hea	+	7.00	98.21	3279	gb_ro:RNRTHRMR
gb_ov:AF16755	7.00	100.87	2206	A76755 Sequence 1 from Patent	-	7.00	98.12	3320	gb_pat:AS8259
gb_pat:AF16765	7.00	100.87	2206	A76760 Sequence 1 from Patent	-	7.00	98.09	3336	gb_pr1:NCR9759
gb_pr1:HS90RTAA	7.00	100.87	2206	A76760 H.sapiens mRNA for 90K	+	7.00	98.05	3359	gb_bal:MTXACS
gb_bal:CVAPETC	7.00	100.84	2215	J03855 C.nostoc rRNA for iron-su	+	7.00	98.04	3363	gb_ro:RNPTRVHRH
gb_pr3:HSM800400	7.00	100.82	2221	AL050074 Homo sapiens mRNA; cD	+	7.00	98.04	3364	gb_ro:SG0053
gb_pat:EO3334	7.00	100.82	2222	E03334 RuBPCase gene of Thoba	-	7.00	97.91	3428	gb_bal:SRPRAEF
gb_ov:AF014368	7.00	100.81	2225	AF014368 Danio rerio transcrip	+	7.00	97.87	3447	gb_pr1:HM80C212
gb_bal:RMHACTS	7.00	100.66	2267	L39938 Sinorhizobium meliloti	+	7.00	97.85	3460	gb_pr3:HSU94333
gb_pr1:HSU18247	7.00	100.66	2275	U18247 Human p67 mRNA, complet	+	7.00	97.80	3485	gb_pr5:AF184971
gb_pr3:HSM802005	7.00	100.63	2285	AL137326 Homo sapiens mRNA; cD	+	7.00	97.68	3546	gb_ro:PSZB8279
gb_pr1:HSGAT1MR	7.00	100.59	2298	A54673 H.sapiens GAT1 mRNA; for	+	7.00	97.60	3591	gb_v1:PSZBETACD
gb_pr1:SCY01148C	7.00	100.59	2299	Z74890 S.cerevisiae chromosome	-	7.00	97.56	3610	gb_in1:BFAMPPIHH
gb_om:TFERB9C	7.00	100.54	2317	D90113 Thioabacillus ferrooxida	+	7.00	97.54	3624	gb_bal:ECTREMER2
gb_v1:MEV133108	7.00	100.52	2323	D26514 Rabbit mRNA for matrix	+	7.00	97.49	3624	gb_pr4:AF132609
gb_v1:MEAFP	7.00	100.35	2384	X00597 Measles virus mRNA for	+	7.00	97.09	3871	gb_ov:PWFGFR
gb_v1:MEAFVUS	7.00	100.35	2384	X00597 Measles virus (Halle) m	+	7.00	97.09	3871	gb_ov:PWFGFR
gb_pr3:HSVAPPEX1	7.00	100.43	2325	U14915 Measles virus, membra	+	7.00	97.42	3687	gb_ro:RATSRH1X
gb_v1:MEAFUS	7.00	100.43	2325	U14915 Measles virus, membra	+	7.00	97.42	3687	gb_ro:RATSRH1X
gb_pr1:ANPDADAGEN	7.00	100.31	2397	M38217 Homo sapiens voltage-ga	+	7.00	96.98	3936	gb_pr5:AF156539
gb_pr1:ANPDADAGEN	7.00	100.31	2397	M38217 Homo sapiens voltage-ga	+	7.00	96.98	3936	gb_pr5:AF156539
gb_bal:ECU45854	7.00	100.38	2400	AF124387 Mus musculus fragile-	-	7.00	96.97	3943	gb_ov:GGV14371
gb_om:MUSNFRP65	7.00	100.24	2424	M61909 Mouse transcription fac	-	7.00	96.87	4000	gb_om:D55682
gb_pr1:ATHAMTB	7.00	100.11	2459	M73467 Arabidopsis thaliana be	+	7.00	96.76	4027	gb_bal:ECU94336
gb_pr3:HSVAPPEX1	7.00	100.04	2494	X98476 H.sapiens VASP gene. 10	+	7.00	96.83	4070	gb_v1:ALUSIN
gb_pr1:ANPDADAGEN	7.00	100.03	2384	X00597 Measles virus (Halle) m	+	7.00	97.09	3871	gb_ov:PWFGFR
gb_pr1:ANPDADAGEN	7.00	100.03	2384	X00597 Measles virus (Halle) m	+	7.00	97.09	3871	gb_ov:PWFGFR
gb_pr1:ANPDADAGEN	7.00	100.03	2384	X00597 Measles virus (Halle) m	+	7.00	97.09	3871	gb_ov:PWFGFR
gb_pr1:ANPDADAGEN	7.00	100.03	2384	X00597 Measles virus (Halle) m	+	7.00	97.09	3871	gb_ov:PWFGFR
gb_bal:ECU45854	7.00	100.38	2400	AF124387 Mus musculus fragile-	-	7.00	96.97	3943	gb_ov:GGV14371
gb_om:MUSNFRP65	7.00	100.24	2424	M61909 Mouse transcription fac	-	7.00	96.87	4000	gb_om:D55682
gb_pr1:ATHAMTB	7.00	100.11	2459	M73467 Arabidopsis thaliana be	+	7.00	96.76	402	

gb_p11:YSCSP014G	-	7.00	94.41	3.3e+03	5778	L46807	Saccharomyces cerevisiae	gb_ba2:AE001817	+	7.00	88.18	7.4e+03	14621	AE001817	Thermotoga maritima
gb_p12:S63539	-	7.00	94.35	3.4e+03	5838	S62539	Insulin receptor subunit	gb_ba1:SC8B7	-	7.00	88.17	7.4e+03	14634	AL031225	Streptomyces coelicolor
gb_p11:EMENIRA	-	7.00	94.34	3.4e+03	5835	MG5900	A.nidulans NIRA protein	gb_ba2:AE001213	-	7.00	88.08	7.5e+03	14838	AE001213	Treponema pallidum
gb_bt6:AC020186	+	7.00	94.28	3.4e+03	5891	AC020186	Drosophila melanogaster	gb_ba2:AS000340	-	7.00	87.93	7.6e+03	15169	AE000340	Escherichia coli
gb_pat:AR064200	+	7.00	94.24	3.4e+03	5924	AR064200	Sequence 6 from patent	gb_ba1:D90801	+	7.00	87.87	7.7e+03	15312	D90801	E.coli genomic DNA
gb_pat:1R9172	-	7.00	94.24	3.4e+03	5924	1R9172	Sequence 6 from patent	gb_bt7:AC010500	+	7.00	87.83	7.7e+03	13396	AC010500	Homo sapiens chromosome
gb_p11:AB033082	-	7.00	94.22	3.4e+03	5938	AB033082	Homo sapiens mRNA for	gb_bt4:AC014953	-	7.00	87.72	7.8e+03	13647	AC014953	Drosophila melanogaster
gb_p11:AB0350631	-	7.00	93.99	3.5e+03	6152	AB0350631	Sequence 1 from patent	gb_ba1:D90884	-	7.00	87.71	7.9e+03	15673	D90884	E.coli genomic DNA
gb_p12:S85963	+	7.00	93.89	3.5e+03	6152	S85963	hhrs-1-rat insulin receptor	gb_bt2:AE000709	-	7.00	87.68	7.9e+03	15737	AE000709	Aquifex aeolicus
gb_p12:HSNFM	+	7.00	93.93	3.6e+03	6236	Y0067	Human gene for neurofil	gb_p11:AP001061	+	7.00	87.64	7.9e+03	15847	AP001061	Homo sapiens genome
gb_p11:D89695	-	7.00	93.89	3.6e+03	6239	D89695	Homo sapiens mRNA for K	gb_pat:AR049868	-	7.00	87.62	8.0e+03	15894	AR049868	Sequence 1 from patent
gb_sy:RAT014131	+	7.00	93.79	3.6e+03	6333	AF174131	Cloning vector PRS411	gb_pat:AE04903	+	7.00	87.62	8.0e+03	13894	AE04903	DNA coding region
gb_sy:RAT014131	+	7.00	93.73	3.6e+03	6389	AF16012	Rattus norvegicus Q-11k	gb_pat:159432	+	7.00	87.62	8.0e+03	15894	159432	Sequence 1 from patent
gb_ro:RNIGPOR	+	7.00	93.17	3.9e+03	6942	Y00826	Rat mRNA for integrin m	gb_vi:MEVMBECDN	+	7.00	87.62	8.0e+03	15894	MEVMBECDN	Sequence 1 from patent
gb_ro:AF113751	+	7.00	93.12	3.9e+03	6998	AF113751	Mus musculus nuclear	gb_bt1:V58435	+	7.00	87.62	8.0e+03	15894	V58435	nucleocapsid protein
gb_p11:AB007931	-	7.00	92.98	4.0e+03	7150	AB007931	Homo sapiens mRNA for	gb_vi:S58435	+	7.00	87.62	8.0e+03	15894	S58435	nucleocapsid protein
gb_sy:AF130854	-	7.00	92.97	4.0e+03	7157	AF130854	Xiphophorus maculatus	gb_ba1:D90747	+	7.00	87.46	8.1e+03	12655	D90747	Escherichia coli gene
gb_sy:AF174132	-	7.00	92.96	4.0e+03	7168	AF174132	Cloning vector PRS421	gb_ro:RATPRPP	+	7.00	87.38	8.2e+03	16453	RATPRPP	Rat proline-rich protein
gb_vi:HS1ATIF1	+	7.00	92.91	4.0e+03	7241	M5621	HSV1 (strain F) alpha-t	gb_ov:AF090337	-	7.00	87.32	8.3e+03	16616	AF090337	Aythya americana coat
gb_ba1:ECMTRETEL	+	7.00	92.81	4.4e+03	8012	Y09025	Enterobacter cloacae DN	gb_pat:AR020969	+	7.00	87.30	8.3e+03	16656	AR020969	Sequence 1 from patent
gb_p12:ACTPOL3GAL	+	7.00	92.17	4.4e+03	8062	L12019	Actinidia deliciosa var	gb_pat:AR043384	+	7.00	87.30	8.3e+03	16656	AR043384	Sequence 1 from patent
gb_sy:AF174133	+	7.00	92.07	4.5e+03	8182	AF174133	Cloning vector PRS421	gb_pat:AR062299	+	7.00	87.30	8.3e+03	16656	AR062299	Sequence 1 from patent
gb_p11:AE0003289	-	7.00	91.47	4.9e+03	8953	AE0003289	Hepatitis B virus C	gb_ba1:D90885	-	7.00	87.11	8.5e+03	17134	D90885	E.coli genomic DNA
gb_ro:MMU217153	+	7.00	91.39	4.9e+03	9060	Y09721	Mycoplasma genitalium s	gb_in1:CEUT20H4	-	7.00	87.10	8.5e+03	17172	CEUT20H4	Sequence 1 from patent
gb_vi:1A16568	+	7.00	91.36												

gb_pr3:HSJ876B10 - 7.00 73.11 5.0e+04 138056 ! AL117352 Human DNA sequence
 gb_pr4:AC004908 - 7.00 73.10 5.0e+04 138251 ! AC004908 Homo sapiens PAC cl
 gb_pr4:AC005280 - 7.00 73.09 5.1e+04 138383 ! AC005280 Homo sapiens clone
 gb_pr4:AC002879 - 7.00 73.09 5.1e+04 138445 ! AC002879 Homo sapiens, *** S
 gb_pr4:HSJ5339224 - 7.00 73.08 5.1e+04 138630 ! AL121920 Homo sapiens chrom
 gb_pr4:AC012388 - 7.00 73.07 5.1e+04 138758 ! AC012388 Drosophila melanoga
 gb_pr4:AC010668 - 7.00 73.06 5.1e+04 139051 ! AC010668 Drosophila melanoga
 gb_pr4:AL137062 - 7.00 73.05 5.1e+04 139327 ! AL137062 Homo sapiens chrom
 gb_pr4:HS6802 - 7.00 73.04 5.1e+04 139389 ! 282215 Human DNA sequence fr
 gb_pr4:D90905 - 7.00 73.04 5.1e+04 139467 ! D90905 Synecocystis sp. PC
 gb_pr4:AC008025 - 7.00 73.04 5.1e+04 139503 ! AC008025 Homo sapiens clone
 gb_pr4:AC019315 - 7.00 73.03 5.1e+04 139700 ! AC019315 Homo sapiens chrom
 seq_name: gb_ba2:AE000213
 seq_documentation_block: 10959 bp DNA BCT 12-NOV-1998
 LOCUS AE000213
 DEFINITION Escherichia coli K-12 MG1655 section 103 of 400 of the complete
 genome.
 ACCESSION AE000213 U00096
 VERSION AE000213.1 GI:1787371
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 (bases 1 to 10959)
 Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.
 The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)
 97426617
 2 (bases 1 to 10959)
 Blattner, F.R.
 Direct Submission
 Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 3 (bases 1 to 10959)
 Blattner, F.R.
 Direct Submission
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 4 (bases 1 to 10959)
 Plunkett, G. III.
 Direct Submission
 Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K-12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: mark@cam.ac.uk). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K-12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications:

and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES

source
 1. .10959
 /organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 13. .40
 /note="factor Sigma70; predicted +1 start at 1184959"
 119. .136
 /note="central position to predicted promoter:82"
 /bound_molecy="PhoB predicted site"
 155. .1381
 /gene="pept"
 /note="bll127"
 155. .1381
 /gene="pept"
 /EC_number="3.4.11.1"
 /function="putative enzyme; Not classified"
 /note="0408; 100 pct identical to 42 aa fragment
 PEPT_ECOLISW: P29745; 93 pct identical to 407 residues of
 409 aa PEPT_SALTY SW: P26311"
 /codon_start=1
 /transl_table=11
 /product="putative peptidase T"
 /protein_id="AAC74211.1"
 /db_xref="GI:1787372"
 /db_xref="PID:gl1787372"
 /translation="MDKLLERLVNLSLTQSKAGVQVPSTEGQWKLHLLKEOLEE
 MGLNLTSEKGLMATLPANVPDIPAGIFSHVDTSDCSGKNVQIVENYRGGD
 IALGDDEVLSPPNFPVHQLLGQTLITDGLGADDAKAGAEIMTAIVQQRKI
 PGDIRVAFPTDEEVGKGAHFDVDAFDARWATYDGGGVELEFENFAASVNRKIV
 GNIHPFGTAGVYVNAISLAIRHAEVPADESPTEGEGYHLASMKGTVERADMH
 YVVPDRKQFEARKKMEIAKVKVGLHPDCYIELVIEDSYNNRKKVVEHPHILD
 IAQAMRDCDIEPELKPINGTGDAQLSFMGLPCPNLFTGGYNGHKHEFVTEGMEK
 AVQVIVRIAEILTAQRK"
 complement(1430..2560)
 /gene="ycfD"
 /note="ycfD"
 complement(1430..2560)
 /gene="ycfD"
 /function="orf; Unknown"
 /note="f376; 100 pct identical to fragment YCFD_ECOLI
 SW:P27431 but has 3 additional N-terminal aa and 93
 additional C-terminal aa"
 /codon_start=1
 /transl_table=11
 /product="orf; hypothetical protein"
 /protein_id="AAC74212.1"
 /db_xref="GI:1787373"
 /db_xref="PID:gl1787373"
 /translation="MLNMEYQLTLNWPDLERHWKRPVYLRKGFNNFIDPISPDELA
 GRAMESEVSRVLSHODGKQVSHGPFESYDHLGETNWSLLVQAVNHWHEPTAALMRP
 FRELPWRIDDLMISSVPVGGVGHLDQYDFIIQGTGRRRRYVGEKQMKQHCMP
 DLLQVDFPRAIDDEELPGLDILYPPGPHGVALENAMNYSVGPAPNTRRLISGFA
 DYVQLRELGNTISDPDPPRAHPADVLPQEMDKREMMELELINQPEHFKQNGFERIS
 QSHELDIAPPEPPYQPDPIYDALKQGEVLVRLGRLVIRIGDVIYANGKERIDSPHRP
 ALDALASNTALAEFGDALEDFSLAALVNSGYWFFEG"
 complement(2583..2613)
 /note="factor sigma70; predicted +1 start at 1187488"
 complement(2627..4087)
 /gene="phoQ"
 complement(2627..4087)
 /note="bll29"
 complement(2627..4087)
 /gene="phoQ"
 /EC_number="2.7.3."
 /function="enzyme; Global regulatory functions"

```

/note="f486: 99 pct identical to PHOQ_ECOLI SW: P23837"
/codon_start=1
/transl_table=11
/product="sensor protein PhoQ"
/protein_id="AAC74213.1"
/db_xref="GI:1787374"
/db_xref="PID:1787374"
/transl_table="MKKLLRFPPLSLVRFLATAAUVLVLSLAYGNVALIGYSVSF
DKITFRLLGSENLYTIAKNNKHLVLPENIQKPTMTLIYDENGQLLUAQRDY
PVMKXIOPDWLKSNGFHEIADVNTSLLSGDHSIQOQOEVEDDDDDAEHTSVA
VNWYPATSPMLKLTIVVDVTIPVELKSSYMWMSFYIVLSANLLVLPVLLVAWWSL
RPIEALKEVRELLREHRELLNPATRELTSLVRLNLLKRSERYDKYRTLDLT
HSLKPLAVLOSTLASLSEKNSVSDAEPVLEQISRSQQIGYVYLHRASMRGGTLLS
RELHPVAPLLDNLTSALKNVYORKGVNISLDSPEISFVGEQNDVEVWGVNVLNACK
YCLEFVESARTDHLXIVVEDDGPGLPSAREVIFDRGQRVDTLRPGQGVGLAVAR
EITEQEYCKIVAGESMLGARMEVIFGRQHSAPKDE"
/complement(4087..4758)
/gene="phop"
/note="b1130"
/complement(4087..4758)
/gene="phop"
/function="regulator; Global regulatory functions"
/note="f223: 100 pct identical to PHOP_ECOLI SW: P23836"
/codon_start=1
/transl_table=11
/product="transcriptional regulatory protein"
/protein_id="AAC74214.1"
/db_xref="GI:1787375"
/db_xref="PID:1787375"
/transl_table="NRVLVVEDNALLRHHLKVOIQDAGHVDVDAEADAYLNEHI
PDIAIVDLGDEPGLSIRWRSNDVSLPILVLTARESDWKDQVLESLAGLDPVTRK
PHIEVYMRQMALRRNGSLASQVSLPFPFVDLSRRELSINDEVIKLTAREYTIMET
LIRNGKVVKSQSLMLQPLDELRESHTIDVLMGRKKIKQAIQVPEVITTVRGQY
LFELR"
/complement(4257..4287)
/gene="phop"
/note="factor Sigma70; predicted +1 start at 1189162"
/complement(4927..5297)
/gene="purB"
/note="b1131"
/complement(4927..6297)
/gene="purB"
/EC_number="4.3.2"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="f456: 99 pct identical to PURB_ECOLI SW: P25739"
/codon_start=1
/transl_table=11
/product="adenylosuccinate lyase"
/protein_id="AAC74215.1"
/db_xref="GI:1787376"
/db_xref="PID:1787376"
/transl_table="MELSLTAVSPVDGRYDKVSALRGIFSEYGLLKFRRVQVEVRWL
KLAHAATKEVPAFAADAIGYLDIAIVASFSEDAARIKTERTNHDVKAYEYELKE
KVAEIPELHVSFEETHEACTSDINLHSLMLKTADEVLVYWRQLIDGKDLAVO
YRDIPLSTHQCPATPSTIGEMANVAYMERQYRQLNOVELIKINGAVGNVNAHI
AAIPEVDHQSFESEVTSUGIOWNPITQIEPHDYIAELFCVAFNLIIDFDSDVW
GYALNHFKQKTAGEISGSTPHKVPNIPDFENSEGNLGSNAQLHSLAKLPVSRWQ
RLDTSTVLRNLGVGIGVALLIAYQSTLAGVSKLVNRYRDLHLDHNEVLAEPITQV
MRYGTEKPYEKLKELTRGKRVDAGMKGQFIDGLALPEEKARLKAMTPANYIGRAIT
MYDELK"
/complement(6301..6942)
/gene="ycfC"
/note="b1132"
/complement(6301..6942)
/gene="ycfC"
/function="orf; Unknown"
/note="f213: 99 pct identical to YCF_C_ECOLI SW: P25746"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"

```

alignment_scores: 52.00 Length: 52

```

Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AE000213/rev ..
Align seg 1/1 to reverse of: AE000213 from: 1 to: 10959

151 TrpLeuTrpProArgAsnProProlleArgGluArgLysSerIleProTh 167
|||||
8818 TGGTTATGGCGGGGAATCCACCAATTCGTGAACGCAAAAGTATCCAC 8769
|||||
167 rSerTrpLeuLysIleThrLeuTrpGluGlyArgAsnArgGlnValArgA 184
|||||
8768 CAGCTGGCTGAAGATCACCTATATGAAGGACGTAATCGCCAGGTGGCC 8719
|||||
184 rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla 200
|||||
8718 GCATGACCGCCCATGTGGCTTCCCCACGCTGCGACTGATTCGCTATGCG 8669
|||||
201 MetGly 202
8668 ATGGGT 8663

seq_name: gb_bal:D90748

seq_documentation_block:
LOCUS D90748 15007 bp DNA BCT 07-FEB-1999
DEFINITION Escherichia coli genomic DNA. (25.6 - 25.9 min).
ACCESSION D90748 AB001340
VERSION D90748.1 GI:1651553
KEYWORDS Complete and shotgun sequencing; potB; potA; pepT; phoQ; phoP;
purB; ycfC; ycfB; lcdA; lcdE; lit.
SOURCE Escherichia coli (strain:K12) DNA, clone:Kohara clone #239.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 15007)
AUTHORS Mori,H.
DIRECT SUBMISSION
SUBMITTED (29-JUL-1996) to the DDBJ/EMBL/GenBank databases.
Hirotsada Mori, NARA Institute of Science and Technology, Res. 6
Edu. Center for Genetic Info.: 8916-5 Takayama Ikoma, Nara 630-01,
Japan [E-mail:hmori@gtc.aisc-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669]
REFERENCE 2 (sites)
AUTHORS Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
Yamamoto,Y. and Yano,M.
The systematic sequencing of the Escherichia coli genome in Japan
Unpublished (1996)
REFERENCE 3 (sites)
AUTHORS Oshima,T., Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
Nishio,Y., Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K.,
Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)
Collaboration Information:
Project: 97061202
The Japan E.coli genome DNA sequencing project
Group: The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,

```

Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.

Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishori@nibb.ac.jp
 Information operator:
 Name: Hirotada Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmorigtc@aist-nara.ac.jp
 URL:
 http://bsw3.aist-nara.ac.jp

Location/Qualifiers
 1. 15007
 /organism="Escherichia coli"
 /strain="kl2"
 /db_xref="taxon:562"
 /clone="Kohara clone #239"
 /map="25.6-25.9 min"
 /note="Nucleotide position 1185468-1200474 from the initiation site of *thrA* (0 min.); This clone is from Kohara lambda miniset library."
 complement(1. .678)
 /gene="potB"
 complement(<1. .678)
 /gene="potB"
 /note="ORF ID:0238#14; similar to SwissProt Accession Number P23860"
 /codon_start=1
 /transl_table=11
 /product="Spermidine/putrescine transport system permease protein PotB."
 /protein_id="BAA35947.1"
 /db_xref="GI:1651554"
 /translation="MVTIVGVLFVFLPMLIGSFELTRDASFKVWVFLDNYT RLDPYFVLLSLNATACINLVGFPAWFLAKLHKVRPLLLFLLIVPFTN SLIDYGLKFLSTKGLNFEFLMGVIDPIRMFTSPSAVIGLVLLIPFVWMPY SSIBKLPLEARDLGASKLQTFIRIIIPLMPTGIAGCLLVMLPAMGLFYVSDLM GGAKNLLI"
 complement(692. .1828)
 /gene="potA"
 complement(692. .1828)
 /gene="potA"
 /note="ORF ID:0239#1; similar to PIR Accession Number A40840"
 /codon_start=1
 /transl_table=11
 /product="Spermidine/putrescine transport protein A"
 /protein_id="BAA35948.1"
 /db_xref="GI:1651555"
 /translation="MGOSKLNKOPSSLPVLQAGIRKCFDGEVQIPQLDITNNHG FLTLGPGGGKTVTLRLIAGLETVDGSRIMLNEIDITHVPAENRYVNTVFQSYALFP HMTVFENATGLRMOKTPAAEITPRVMEALRMVQLETFAQRKPHLSGGQOQVAIAR AVNKPRLLLDESLLADYKLRKQMONELKALQKIGITFVFTHDQEAALPMSDEI VVWDGRIEDGCPREIYEPKNIIFYAGFGEINMENVATVIERLDEORVRANVEGREC NIYNVAFEGQLKLVLRPELDLVEINDNNAEGLIGVYRERNYKGMTLESVVELE NGRMVSEFFNDDPDDFDSLQDKMAINWVESWEVVIADDEHK"
 2078. .3304
 /gene="pept"
 2078. .3304
 /gene="pept"
 /note="ORF ID:0239#2; similar to SwissProt Accession Number P29745"
 /codon_start=1
 /transl_table=11
 /product="Peptidase T (EC 3.4.11.-) (aminotripeptidase)

(tripeptidase)."
 /protein_id="BAA35949.1"
 /db_xref="GI:1651556"
 /translation="MDKLLERFLNVSLDTQSKAGVQVPSTEGQHKLLHLKEOLEE MGLINTLSEKGTLMATLPANVPGDIPAGTISHVDTSPDCSKNNVQIENVYRGD IALGIDGVLSPVNFVHQLLQTLITTDGKTLGAGDADKAGIAETMTALVLOQRKI PHGDIRVATPDVEVGKAGKHFVDADFARWAYTVDDGGVGELEFENFAASVNIKIV GNNVHPTAGVMVNAISLAARITAEVPADESPETEGVEGFYHLASMGKTVERADMH YIIRDTRQKFEARKKMEIAKKVGKGLHPDCYIELVIEDSYNNREKVVVERPHILO IAQAMRDCDIEPELPIRGDTGAQLSFGNGLPCPNLFTGGYNYHGKHEFVILEMEK AVQVIRIAELTAQRK"
 complement(3353. .6010)
 /gene="phoQ"
 complement(3353. .4474)
 /gene="phoQ"
 /note="ORF ID:0239#3; similar to PIR Accession Number D41966"
 /codon_start=1
 /transl_table=11
 /product="Orf 2 downstream of phoQ."
 /protein_id="BAA35950.1"
 /db_xref="GI:4062695"
 /translation="MEYQTLNWPDELERHWQRPVVKRGFNNFIDPISDELACIA MESEVSRVSHQDGRWQVSHGPFESIDHGETNWSLLVQAVNHWHEPTAALMRPFE LPDRIDDLMSFSVPGGVGPHLDQYDFIIGTGRRRVRKQKMHQCHPHDPL QVDPFEAIIDEELEPGDILYIPGPFEGYALENANNYSVGRAPKTRELISGFADYV LORELGNTYSDPDVPPRAHPADVLPOEMDKREMMLINQPEHKEKQWGFISQSR HELDIAPPPEYQPDIEYDALKQGEVLVRLGGLVRLIGDDVYANGEKIDSPHRPALD ALASNTALTAENEGDALEDPSEFLAALVNSGYWFEFG"
 complement(4550. .6010)
 /gene="phoQ"
 /note="ORF ID:0239#4; similar to PIR Accession Number B41966"
 /codon_start=1
 /transl_table=11
 /product="Virulence membrane protein phoQ."
 /protein_id="BAA35951.1"
 /db_xref="GI:1651557"
 /translation="MKLLRFLFPLSLRVRLFLATAAVVLVLSLAYGMVALIGYSVSF DKTTPRLRGESNLFYTLAKWENKHLVELPENIDKQSPMTLIYDENGQLLWQARDV PMLKIQPDLKSGFHEIADVNDTSLISGDHSIQOQLVEVREDDDAETHSYA VNYPATSMPLTIVVDTIPVELKSSYMWVSFVIYLSANLLVPLLVNAAWMSL RPTALAKEVELEHNELNPATRELTSILVNLNRLKSEREYDKYRTLTDLT HSLKTPAVLQTLRSLSRSEKSVSDAEPVMEQISRIQQIYGLHRASMRGTLIS RELHPVAPLIDNLTSALNKVYQKGNISLIDISPEISFVGEQDNFVEMGNVLDNACK YCLEFVEISARQTDHELYIVVEDDGPICPLSKREVIFDRGQVDTLRPGQGVGLAVAR EITEQYEGKIVAGESMIGGARMEVIFGRQHSAPKDE"
 complement(6010. .6681)
 /gene="phoP"
 complement(6010. .6681)
 /note="ORF ID:0240#1; similar to SwissProt Accession Number P23836"
 /gene="phoP"
 /codon_start=1
 /transl_table=11
 /product="Transcriptional regulatory protein PhoP."
 /protein_id="BAA35952.1"
 /db_xref="GI:1651558"
 /translation="MRVLVEENALRLHKLKQIDAGHQVDDAEADKADYVLEHI PDIIVDLGLPDEGLSLRRRNSDVSLPILVLTARESQKVEVLSCADYVTKP FHIEVWAKQALRRNSGLASQVLSLPPFQDLSRLRELSINDEVIKLTAFETIMET LIRNKKVVSQKSLMLQIPLPYDAELRESHTIDVLMRLKRKKIQAYPQEVITTVRGOGY LPELR"
 complement(6850. .8220)
 /gene="purB"
 complement(6850. .8220)
 /gene="purB"
 /note="ORF ID:0240#2; similar to PIR Accession Number S19212"
 /codon_start=1
 /transl_table=11
 /product="Adenylosuccinate lyase (EC 4.3.2.2)"
 /protein_id="BAA35953.1"


```
/gene="purb"
/notes="ORF_ID:0240#2; similar to PIR Accession Number
S19212"
/codon_start=1
/transl_table=11
/product="Adenylosuccinate lyase (EC 4.3.2.2)"
/protein_id="BAA35962.1"
/db_xref="GI:1651565"
/translation="MELSSLTAVSPVDGRYGDVSKVALRGIFSEYGLKFRVQVEVRWL
QKLAHAAKVPFAADAIYGLDAIVASFSEEDAARIKTIERTNHDHVAEYFLKE
KVAETPELHAYSETHFACTSEDINNLSHMLKTAARDEVILPYWRQIDGKDLAVQ
YRDLPLSLRGTQGPATPTIGKEMANVAYRMEROYRQNLQVEIILKINGAVGNTNAHI
RAYPEVDHOFSEEVTSLSLQWNPYTQIEPHOYIAELFCVAREFTILLDFEDVW
GYIALNHFKQTIAGEICSSMTPHKVNFIDENSEGNLGLSNVQLHLSKLPVSRWQ
RDLTSTVRLNGLVGIGVALLAYOSTLKGVSKEVLRNDRHLLDELHDHNEVLAEPYQV
MRRYGIEPKLKELTGRKVRDAEGMKQFIDGLALPPEERKARKAMTPANYIGRAIT
MVDELK"
/complement(2541..3182)
/gene="ycfc"
/complement(2541..3182)
/notes="ORF_ID:0240#3; similar to PIR Accession Number
S19211"
/codon_start=1
/transl_table=11
/product="Hypothetical protein 23."
/protein_id="BAA35963.1"
/db_xref="GI:4062714"
/translation="MAKNYDITLALAGICQARLVQQLAHQGHCDADALHVSLSNII
DMNPSSTLAVFGSEANLRVGLTLLGVNASSRQGLNAELRYTLISLWVERKLSSA
KGALDTNRRINGLRQEDLOSETLMSAMAAIYDVISPGRIGVQTSQPAVLQS
PQVAKVLRATLIGIRAVLAWHVGGRQLQMFSENRLTTQAKILAHITPEL"
/complement(3218..4324)
/gene="ycfb"
/complement(3218..4324)
/notes="ORF_ID:0240#4; similar to SwissProt Accession
Number P25745"
/codon_start=1
/transl_table=11
/product="Hypothetical protein in purb 5'region (orf-15)"
/protein_id="BAA35964.1"
/db_xref="GI:4062715"
/translation="MSETRAKVIVMGSGVDSSYSAWLLQQQGVQVBLEFMKNWEEDD
GEETCAARDADAAQVCDKGLIELHTVNFAAEYWDNVFELFAEYKAGRTPNPDILC
NKEIKFRAFLFEADGADYATGHYVVRADVDGKSLRLGLDSNKDQSYFLTLSH
EQIAQSLFPVGELEKPOVKYIAEDGLVLTAKKDKSTGICFGERKREFLGRYLPVQ
GKIITVDGEIGEHQGLMYHTLGQRKGLIGGTGTEEPWYVVDKDVNNILVPAQG
HEHPLMSVGLIAQLHWDVREPFTGMRCVTKYRQTDIPICVKALEDDRIEVIDF
EPVAATVPGQSAFYNGEVCVGLGGLIEQRLEPV"
/complement(4378..4839)
/notes="ORF_ID:0240#5; similar to PIR Accession Number
S64074"
/codon_start=1
/transl_table=11
/product="Hypothetical protein YGL067w"
/protein_id="BAA35965.1"
/db_xref="GI:4062716"
/translation="MFKPHVTVACVHAEKFLVVEETINGKALWNPAGHLEADETL
VEAAARELWETGISAQPHQHWIAPDKTPTFLRFLFAIELEQICPTOPHSDID
CCRWYSAEIILQASNLRSPLVAESRCVQSQRYPLEMGFNWFFKGV"
/complement(4849..5502)
/notes="ORF_ID:0240#6; similar to PIR Accession Number
164156"
/codon_start=1
/transl_table=11
/product="Hypothetical protein HI0694"
/protein_id="BAA35966.1"
/db_xref="GI:4062717"
/translation="MRQFIISENTMOKTSFRNHQKVFSSQSRSTRKPNQPTRVILF
NKPYDLPQFDAGRTKLEFIPVQVYAAAGRLDRDSEGLLVLTNNGALQARLTQPC
KRTGIYVQVEGIPQDALEALRNGVTINDGPTLPAGAEVLDPEFALWPRNPPIR
```

```
KSPTSWLKLTLYEGRNQYRRMTAHVGFPTLRLIRYAMGDYSLDNLANGEWREVTD"
5674..6924
/gene="icd"
5674..6924
/gene="icd"
/notes="ORF_ID:0240#7; similar to SwissProt Accession
Number P08200"
/codon_start=1
/transl_table=11
/product="Isocitrate dehydrogenase (NADP) (EC 1.1.1.42)
(Oxalosuccinate decarboxylase) (icdh) (NADP+-specific icdh)
(icdp)"
/protein_id="BAA35967.1"
/db_xref="GI:1651566"
/translation="MESKVVVPAQGGKITLQNGKLVNPNPIIPIEGDIGVDVTPA
MLKVVDAAVKAYKGERKISWMEIYTGKSTQVYQDQWLPAAETDLIRYRVAKIDY
LTPVGGGIRSLNVALRQELDLVLCRLPVRVYOGTSPVKHPELTDNVIREFNSEDIY
AGLEMKADSADAQVVKFLEEMGVKKIRFEHCGGIGIKPCSEGGTRLYRAALEYAI
ANDRDSVTLVHKNIMKFTGAFDMGQLAREFEFGGELIDGGFWLKVKNPNTGKEIV"
alignment_scores:
Quality: 52.00 Length: 52
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-252-691-7056 x D90749/rev ..
Align seg 1/1 to reverse of: D90749 from: 1 to: 20284
151 TTPLeuTTPProArgAsnProIleArgGluArgLysSerIleProTh 167
|||||
5058 TGGTATGGCCGGGAATCCACCAATTCGTGAACGAAAGTATTCGCCAC 5009
167 rSerTrpLeuLysIleThrLeuTyGluGlyArgAsnArgGlnValArgA 184
|||||
5008 CAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCCGACAGTGGGCC 4959
184 rGMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyAla 200
|||||
4958 GCATGACCGCCCATGTGGCTTCCCGACGCTGCAGCTGCTATGCGTATGGG 4909
201 MetGly 202
|||||
4908 ATGGGT 4903
seq_name: gb_htg7:AC022161
seq_documentation_block:
LOCUS AC022161 180993 bp DNA HTG 26-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-273P11, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC022161
VERSION AC022161.1 GI:6758614
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 180993)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 180993)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
```

* NOTE: This record contains 113 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 511: contig of 511 bp in length
* gap of unknown length
* 512 1507: contig of 996 bp in length
* gap of unknown length
* 1508 2279: contig of 772 bp in length
* gap of unknown length
* 2280 2614: contig of 335 bp in length
* gap of unknown length
* 2615 3450: contig of 836 bp in length
* gap of unknown length
* 3451 3914: contig of 464 bp in length
* gap of unknown length
* 3915 4858: contig of 944 bp in length
* gap of unknown length
* 4859 5380: contig of 522 bp in length
* gap of unknown length
* 5381 6247: contig of 867 bp in length
* gap of unknown length
* 6248 7203: contig of 956 bp in length
* gap of unknown length
* 7204 7920: contig of 717 bp in length
* gap of unknown length
* 7921 8924: contig of 1004 bp in length
* gap of unknown length
* 8925 9674: contig of 750 bp in length
* gap of unknown length
* 9675 11049: contig of 1375 bp in length
* gap of unknown length
* 11050 11834: contig of 785 bp in length
* gap of unknown length
* 11835 12774: contig of 940 bp in length
* gap of unknown length
* 12775 13555: contig of 781 bp in length
* gap of unknown length
* 13556 14538: contig of 983 bp in length
* gap of unknown length
* 14539 15398: contig of 860 bp in length
* gap of unknown length
* 15399 15696: contig of 298 bp in length
* gap of unknown length
* 15697 16675: contig of 979 bp in length
* gap of unknown length
* 16676 17597: contig of 922 bp in length
* gap of unknown length
* 17598 18969: contig of 1372 bp in length
* gap of unknown length
* 18970 19300: contig of 331 bp in length
* gap of unknown length
* 19301 19380: contig of 80 bp in length
* gap of unknown length
* 19381 20457: contig of 1077 bp in length
* gap of unknown length
* 20458 21187: contig of 730 bp in length
* gap of unknown length
* 21188 21435: contig of 248 bp in length
* gap of unknown length
* 21436 22621: contig of 1186 bp in length
* gap of unknown length
* 22622 23209: contig of 588 bp in length
* gap of unknown length
* 23210 23994: contig of 785 bp in length
* gap of unknown length

* 23995 24643: contig of 649 bp in length
* gap of unknown length
* 24644 25421: contig of 778 bp in length
* gap of unknown length
* 25422 25826: contig of 405 bp in length
* gap of unknown length
* 25827 27532: contig of 1706 bp in length
* gap of unknown length
* 27533 28937: contig of 1405 bp in length
* gap of unknown length
* 28938 30011: contig of 1074 bp in length
* gap of unknown length
* 30012 30574: contig of 563 bp in length
* gap of unknown length
* 30575 31545: contig of 971 bp in length
* gap of unknown length
* 31546 32314: contig of 769 bp in length
* gap of unknown length
* 32315 33010: contig of 696 bp in length
* gap of unknown length
* 33011 33654: contig of 644 bp in length
* gap of unknown length
* 33655 34705: contig of 1051 bp in length
* gap of unknown length
* 34706 35755: contig of 1050 bp in length
* gap of unknown length
* 35756 36672: contig of 917 bp in length
* gap of unknown length
* 36673 37904: contig of 1232 bp in length
* gap of unknown length
* 37905 38457: contig of 553 bp in length
* gap of unknown length
* 38458 39301: contig of 844 bp in length
* gap of unknown length
* 39302 40257: contig of 956 bp in length
* gap of unknown length
* 40258 41691: contig of 1434 bp in length
* gap of unknown length
* 41692 43227: contig of 1536 bp in length⁹
* gap of unknown length
* 43228 44416: contig of 1189 bp in length
* gap of unknown length
* 44417 45206: contig of 790 bp in length
* gap of unknown length
* 45207 46223: contig of 1017 bp in length
* gap of unknown length
* 46224 47104: contig of 881 bp in length
* gap of unknown length
* 47105 49150: contig of 2046 bp in length
* gap of unknown length
* 49151 50227: contig of 1077 bp in length
* gap of unknown length
* 50228 51990: contig of 1763 bp in length
* gap of unknown length
* 51991 52651: contig of 661 bp in length
* gap of unknown length
* 52652 54157: contig of 1506 bp in length
* gap of unknown length
* 54158 55684: contig of 1527 bp in length
* gap of unknown length
* 55685 56936: contig of 1252 bp in length
* gap of unknown length
* 56937 57993: contig of 1057 bp in length
* gap of unknown length
* 57994 58875: contig of 882 bp in length
* gap of unknown length
* 58876 60147: contig of 1272 bp in length
* gap of unknown length
* 60148 61314: contig of 1167 bp in length
* gap of unknown length
* 61315 61502: contig of 188 bp in length
* gap of unknown length
* 61503 62298: contig of 796 bp in length

```
* * gap of unknown length
* * 62299 63529: contig of 1231 bp in length
* * gap of unknown length
* * 63530 65376: contig of 1847 bp in length
* * gap of unknown length
* * 65377 65461: contig of 85 bp in length
* * gap of unknown length
* * 65462 66798: contig of 1337 bp in length
* * gap of unknown length
* * 66799 68065: contig of 1267 bp in length
* * gap of unknown length
* * 68066 69742: contig of 1677 bp in length
* * gap of unknown length
* * 69743 71264: contig of 1522 bp in length
* * gap of unknown length
* * 71265 72180: contig of 916 bp in length
* * gap of unknown length
* * 72181 74061: contig of 1881 bp in length
* * gap of unknown length
* * 74062 75047: contig of 986 bp in length
* * gap of unknown length
* * 75048 76521: contig of 1474 bp in length
* * gap of unknown length
* * 76522 78628: contig of 2107 bp in length
* * gap of unknown length
* * 78629 80347: contig of 1719 bp in length
* * gap of unknown length
* * 80348 82199: contig of 1852 bp in length
* * gap of unknown length
* * 82200 83804: contig of 1605 bp in length
```

```
alignment_scores:
  Quality: 42.00      Length: 42
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment block:

US-09-252-691-7056 x AC022161

Align seg 1/1 to: AC022161 from: 1 to: 180993

```
151 TrpLeuTyrProArgAsnProIleArgGluArgLysSerIleProth 167
|||||
32188 TGCTATGCGCGGAATACCAACCAATTCGTGACGCAAAAGATTCCAC 32237
|||||
167 rSerTyrLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
|||||
32238 CAGCTGGCTGAAGATCACCTTATATGAGGACGTAATCGCCAGGTGCGCC 32287
|||||
184 rGMetThrAlaHisValGlyPhePro 192
|||||
32288 GCATGACCGCCCATGTGTGCTTCCCC 32313
```

seq_name: gb_hlg6:AC008911

seq_documentation block:

```
LOCUS AC008911 226805 bp DNA HTG 14-JAN-2000
DEFINITION Homo sapiens chromosome 5 clone CITB-HL_2268J5, WORKING DRAFT
SEQUENCE, 95 unknown pieces.
ACCESSION AC008911
VERSION AC008911.2 GI:6693307
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 226805)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 226805)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
```

JOURNAL

COMMENT

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 14, 2000 this sequence version replaced gi:5686162.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>
-----Summary Statistics
Consensus quality: 120029 bases at least Q40
Consensus quality: 163743 bases at least Q30
Consensus quality: 177506 bases at least Q20
Estimated insert size: 226805; sum-of-contigs estimation
Estimated insert size: 164170; agarose-fp estimation
Quality coverage: 4.03x in Q20 bases; agarose-fp estimation
Quality coverage: 2.91x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently consists of 95 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 1018: contig of 1018 bp in length
* * gap of unknown length
* * 1019 2070: contig of 1052 bp in length
* * gap of unknown length
* * 2071 3108: contig of 1038 bp in length
* * gap of unknown length
* * 3109 4211: contig of 1103 bp in length
* * gap of unknown length
* * 4212 5226: contig of 1015 bp in length
* * gap of unknown length
* * 5227 6336: contig of 1110 bp in length
* * gap of unknown length
* * 6337 7420: contig of 1084 bp in length
* * gap of unknown length
* * 7421 8436: contig of 1016 bp in length
* * gap of unknown length
* * 8437 9514: contig of 1078 bp in length
* * gap of unknown length
* * 9515 11442: contig of 1928 bp in length
* * gap of unknown length
* * 11443 12908: contig of 1466 bp in length
* * gap of unknown length
* * 12909 14243: contig of 1335 bp in length
* * gap of unknown length
* * 14244 15370: contig of 1127 bp in length
* * gap of unknown length
* * 15371 16374: contig of 1004 bp in length
* * gap of unknown length
* * 16375 17484: contig of 1110 bp in length
* * gap of unknown length
* * 17485 18486: contig of 1002 bp in length
* * gap of unknown length
* * 18487 19535: contig of 1049 bp in length
* * gap of unknown length
* * 19536 20747: contig of 1212 bp in length
* * gap of unknown length
* * 20748 21794: contig of 1047 bp in length
* * gap of unknown length
* * 21795 22910: contig of 1116 bp in length
* * gap of unknown length
* * 22911 24510: contig of 1600 bp in length
* * gap of unknown length
* * 24511 26330: contig of 1820 bp in length
* * gap of unknown length
* * 26331 27876: contig of 1546 bp in length
* * gap of unknown length
* * 27877 29050: contig of 1174 bp in length
* * gap of unknown length
* * 29051 30329: contig of 1279 bp in length

```
* * gap of unknown length
* * 30330 31614: contig of 1285 bp in length
* * gap of unknown length
* * 31615 33078: contig of 1464 bp in length
* * gap of unknown length
* * 33079 34288: contig of 1210 bp in length
* * gap of unknown length
* * 34289 35341: contig of 1053 bp in length
* * gap of unknown length
* * 35342 36695: contig of 1354 bp in length
* * gap of unknown length
* * 36696 37729: contig of 1034 bp in length
* * gap of unknown length
* * 37730 39224: contig of 1495 bp in length
* * gap of unknown length
* * 39225 40993: contig of 1169 bp in length
* * gap of unknown length
* * 40994 41828: contig of 1435 bp in length
* * gap of unknown length
* * 41829 43639: contig of 1811 bp in length
* * gap of unknown length
* * 43640 45023: contig of 1384 bp in length
* * gap of unknown length
* * 45024 46157: contig of 1134 bp in length
* * gap of unknown length
* * 46158 47250: contig of 1093 bp in length
* * gap of unknown length
* * 47251 48869: contig of 1619 bp in length
* * gap of unknown length
* * 48870 50138: contig of 1269 bp in length
* * gap of unknown length
* * 50139 51169: contig of 1031 bp in length
* * gap of unknown length
* * 51170 52255: contig of 1086 bp in length
* * gap of unknown length
* * 52256 53454: contig of 1199 bp in length
* * gap of unknown length
* * 53455 55052: contig of 1598 bp in length
* * gap of unknown length
* * 55053 56384: contig of 1332 bp in length
* * gap of unknown length
* * 56385 57426: contig of 1042 bp in length
* * gap of unknown length
* * 57427 58894: contig of 1468 bp in length
* * gap of unknown length
* * 58895 59899: contig of 1005 bp in length
* * gap of unknown length
* * 59900 61097: contig of 1198 bp in length
* * gap of unknown length
* * 61098 62827: contig of 1730 bp in length
* * gap of unknown length
* * 62828 64143: contig of 1316 bp in length
* * gap of unknown length
* * 64144 65320: contig of 1177 bp in length
* * gap of unknown length
* * 65321 66338: contig of 1018 bp in length
* * gap of unknown length
* * 66339 67938: contig of 1600 bp in length
* * gap of unknown length
* * 67939 69534: contig of 1596 bp in length
* * gap of unknown length
* * 69535 70638: contig of 1104 bp in length
* * gap of unknown length
* * 70639 71721: contig of 1083 bp in length
* * gap of unknown length
* * 71722 72737: contig of 1016 bp in length
* * gap of unknown length
* * 72738 74240: contig of 1503 bp in length
* * gap of unknown length
* * 74241 76005: contig of 1765 bp in length
* * gap of unknown length
* * 76006 77033: contig of 1028 bp in length
* * gap of unknown length
```

```
* * 77034 79330: contig of 2297 bp in length
* * gap of unknown length
* * 79331 80402: contig of 1072 bp in length
* * gap of unknown length
* * 80403 82066: contig of 1664 bp in length
* * gap of unknown length
* * 82067 83385: contig of 1319 bp in length
* * gap of unknown length
* * 83386 85506: contig of 2121 bp in length
* * gap of unknown length
* * 85507 87079: contig of 1573 bp in length
* * gap of unknown length
* * 87080 88558: contig of 1479 bp in length
* * gap of unknown length
* * 88559 89799: contig of 1241 bp in length
* * gap of unknown length
* * 89800 92035: contig of 2236 bp in length
* * gap of unknown length
* * 92036 94093: contig of 2058 bp in length
* * gap of unknown length
* * 94094 96588: contig of 2495 bp in length
* * gap of unknown length
* * 96589 98564: contig of 1976 bp in length
* * gap of unknown length
* * 98565 100900: contig of 2336 bp in length
* * gap of unknown length
* * 100901 103171: contig of 2271 bp in length
* * gap of unknown length
* * 103172 105268: contig of 2097 bp in length
* * gap of unknown length
* * 105269 108353: contig of 3585 bp in length
* * gap of unknown length
* * 108354 112992: contig of 4139 bp in length
* * gap of unknown length
* * 112993 117661: contig of 4669 bp in length
* * gap of unknown length
* * 117662 121578: contig of 3917 bp in length
```

alignment_scores:

```
Quality: 28.00 Length: 28
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

```
US-09-252-691-7056 x AC008911/rev ..
```

```
Align seg 1/1 to reverse of: AC008911 from: 1 to: 226805
```

```
151 TrpLeuTrpProArgAsnProPtoIleArgGluArgLysSerIleProTh 167
|||||
50738 TGGTTATGGCGCGGAAATCCACCAATTCGTGACGCAAAAGATATCCAC 50689
```

```
167 rSerTrpLeuLysIleThrLeuTyrGluGlyArg 178
|||||
```

```
50688 CAGCTGCTGAGATCACCCTATATGAGGACGC 50655
```

```
seq_name: gb_htg7:AC009127
```

```
seq_documentation_block:
```

```
LOCUS AC009127 186591 bp DNA HTG 26-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-498D10, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC009127
VERSION AC009127.2 GI:6758895
KEYWORDS HTG; HTGS_PHASE0.
```

```
SOURCE
```

```
ORGANISM
```

```
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 186591)
```

```
AUTHORS DOE Joint Genome Institute.
```

```
TITLE Sequencing of Human Chromosome 16
```

```
JOURNAL Unpublished
```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

2 (bases 1 to 186591)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 26, 2000 this sequence version replaced gi:5685948.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

* NOTE: This record contains 101 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
989: contig of 989 bp in length
gap of unknown length
990 1318: contig of 329 bp in length
gap of unknown length
1319 1843: contig of 525 bp in length
gap of unknown length
1844 2066: contig of 223 bp in length
gap of unknown length
2067 3127: contig of 1061 bp in length
gap of unknown length
3128 3535: contig of 408 bp in length
gap of unknown length
3536 4171: contig of 636 bp in length
gap of unknown length
4172 4797: contig of 626 bp in length
gap of unknown length
4798 5170: contig of 373 bp in length
gap of unknown length
5171 5390: contig of 220 bp in length
gap of unknown length
5391 6021: contig of 631 bp in length
gap of unknown length
6022 6686: contig of 865 bp in length
gap of unknown length
6687 6975: contig of 289 bp in length
gap of unknown length
6976 7331: contig of 356 bp in length
gap of unknown length
7332 7911: contig of 580 bp in length
gap of unknown length
7912 8248: contig of 337 bp in length
gap of unknown length
8249 8844: contig of 596 bp in length
gap of unknown length
8845 9149: contig of 305 bp in length
gap of unknown length
9150 9469: contig of 320 bp in length
gap of unknown length
9470 9911: contig of 442 bp in length
gap of unknown length
9912 10494: contig of 583 bp in length
gap of unknown length
10495 11143: contig of 649 bp in length
gap of unknown length
11144 11762: contig of 619 bp in length
gap of unknown length
11763 12380: contig of 618 bp in length
gap of unknown length
12381 13000: contig of 620 bp in length
gap of unknown length
13001 13576: contig of 576 bp in length
gap of unknown length

13577 14175: contig of 599 bp in length
gap of unknown length
14176 14801: contig of 626 bp in length
gap of unknown length
14802 15373: contig of 572 bp in length
gap of unknown length
15374 15716: contig of 343 bp in length
gap of unknown length
15717 16489: contig of 773 bp in length
gap of unknown length
16490 17111: contig of 622 bp in length
gap of unknown length
17112 17855: contig of 744 bp in length
gap of unknown length
17856 19178: contig of 1323 bp in length
gap of unknown length
19179 20052: contig of 874 bp in length
gap of unknown length
20053 20790: contig of 738 bp in length
gap of unknown length
20791 21464: contig of 674 bp in length
gap of unknown length
21465 22165: contig of 701 bp in length
gap of unknown length
22166 22966: contig of 801 bp in length
gap of unknown length
22967 23458: contig of 492 bp in length
gap of unknown length
23459 24138: contig of 680 bp in length
gap of unknown length
24139 24984: contig of 846 bp in length
gap of unknown length
24985 25642: contig of 658 bp in length
gap of unknown length
25643 26565: contig of 923 bp in length
gap of unknown length
26566 27196: contig of 631 bp in length
gap of unknown length
27197 27815: contig of 619 bp in length
gap of unknown length
27816 27901: contig of 86 bp in length
gap of unknown length
27902 28512: contig of 611 bp in length
gap of unknown length
28513 29881: contig of 1369 bp in length
gap of unknown length
29882 30600: contig of 719 bp in length
gap of unknown length
30601 31204: contig of 604 bp in length
gap of unknown length
31205 32333: contig of 1129 bp in length
gap of unknown length
32334 32911: contig of 578 bp in length
gap of unknown length
32912 33816: contig of 905 bp in length
gap of unknown length
33817 34778: contig of 962 bp in length
gap of unknown length
34779 36264: contig of 1485 bp in length
gap of unknown length
36265 38613: contig of 2349 bp in length
gap of unknown length
38614 39144: contig of 531 bp in length
gap of unknown length
39145 41224: contig of 2080 bp in length
gap of unknown length
41225 41982: contig of 758 bp in length
gap of unknown length
41983 43378: contig of 1396 bp in length
gap of unknown length
43379 44856: contig of 1478 bp in length
gap of unknown length

```
* 44857 46513: contig of 1657 bp in length
* 46514 48278: contig of 1765 bp in length
* 48279 49123: contig of 845 bp in length
* 49124 50185: contig of 1062 bp in length
* 50186 51798: contig of 1613 bp in length
* 51799 52794: contig of 996 bp in length
* 52795 54154: contig of 1360 bp in length
* 54155 56271: contig of 2117 bp in length
* 56272 58993: contig of 2722 bp in length
* 58994 60268: contig of 1275 bp in length
* 60269 62633: contig of 2365 bp in length
* 62634 65099: contig of 2466 bp in length
* 65100 66776: contig of 1677 bp in length
* 66777 68874: contig of 2098 bp in length
* 68875 72266: contig of 3392 bp in length
* 72267 74691: contig of 2425 bp in length
* 74692 77589: contig of 2897 bp in length
* 77589 80107: contig of 2519 bp in length
* 80108 84825: contig of 4718 bp in length
* 84826 87029: contig of 2204 bp in length
* gap of unknown length
```

```
alignment_scores:
  Quality: 26.00      Length: 26
  Ratio: 1.000       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-252-691-7056 x AC009127
```

```
Align seg 1/1 to: AC009127 from: 1 to: 186591
```

```
151 TrpLeuTrpProArgAsnProIleArgGluArgLysSerIleProth 167
|||||
28398 TGGTTATGGCGCGGATCCACCAATTCGTGACGCAAAAGATATCCAC 28447
|||||
167 rSerTrpLeuLysIleHrLeuTyrGlu 176
|||||
28448 CAGCTGGCTTAAGATCACCTTATATGAA 28475
|||||
```

```
seq_name: gb_ba2-U32752
```

```
seq_documentation_block:
LOCUS      U32752      11833 bp      DNA      BCT      29-MAY-1998
DEFINITION Haemophilus influenzae Rd section 67 of 163 of the complete genome.
ACCESSION  U32752 L42023
VERSION     U32752.1 GI:1573692
```

```
KEYWORDS
SOURCE      Haemophilus influenzae Rd.
```

```
ORGANISM    Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
```

```
REFERENCE   1 (bases 1 to 11833)
```

```
AUTHORS     Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
```

```
Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
Merrick,J.M., McInerney,K., Sutton,G.G., Fitzhugh,W., Fields,C.A.,
Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L., Saudek,C.M.,
Georgachen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
Smith,H.O. and Venter,J.C.
Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd
JOURNAL     Science 269 (5223), 496-512 (1995)
MEDLINE     95350630
REFERENCE   2 (bases 1 to 11833)
AUTHORS     Tatusov,R.L., Muehlegan,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
TITLE       Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
JOURNAL     Curt. Biol. 6 (3), 279-291 (1996)
MEDLINE     96398784
REFERENCE   3 (bases 1 to 11833)
AUTHORS     White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
TITLE       Direct Submission
JOURNAL     Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr. Rockville, MD 20850, USA
REFERENCE   4 (bases 1 to 11833)
AUTHORS     White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
TITLE       Direct Submission
JOURNAL     Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr. Rockville, MD 20850, USA
REMARK      The H. influenzae sequence has been updated by R. Fleischmann. New
            database matches have been assigned, product names have been
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5 (bases 1 to 11833)
AUTHORS     White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
TITLE       Direct Submission
JOURNAL     Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr. Rockville, MD 20850, USA
REMARK      The whole genome was shifted by 598 nucleotides for a new start
            On Sep 30, 1996 this sequence version replaced gi:1221393.
FEATURES
            source
            1..11833
            /organism="Haemophilus influenzae Rd"
            /db_xref="taxon:71421"
            145..939
            /gene="HI0690"
            145..939
            /gene="HI0690"
            /note="similar to GB:L19201 SP:L11244 GB:M55990 GB:U13915
            GB:X15054 percent identity: 73.18; identified by sequence
            similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="glycerol uptake facilitator protein (gipf)"
            /protein_id="AAC22350.1"
            /db_xref="GI:1573694"
            /translation="MDKSLKANCIGEEFLGTALLIFPGVCCVAALKVAGASFGLWEISI
            MWMGVALAVATAGLSGAHLNPVATILWKFCDGKKVIFPIISQMGAFFAALV
            YALYRNFDIETVHNIVRGTOESLSLAGTFTSTPHPSISIGGFAVEVITAILMAL
            IMAIADDGNGVPRGPLAPLILIGLIAVIGGANGPLTGFANPARDPFGFFAYLAGWG
            EALATGGGREIPYFIVMPVAPVLGALAGAWLYKKAIGGNLPCNCGE"
            960..2471
            /gene="HI0691"
            960..2471
            /gene="HI0691"
            /note="similar to GB:L19201 SP:P08859 GB:M18393 GB:M55990
            GB:X15054 percent identity: 76.89; identified by sequence
            similarity; putative"
            /codon_start=1
            /transl_table=11
```


TITLE Characterization of the functional involvement of adeoxyxylulose 5-phosphate reductoisomerase gene harbouring locus of the *Synechococcus leopoliensis* genome in isoprenoid biosynthesis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 5476)

AUTHORS Zimmer, W.

TITLE Direct Submission

JOURNAL Submitted (29-OCT-1999) Zimmer W., Atmosphärische Umweltforschung, Fraunhofer Institut, Kreuzackbahnstrasse 19, Garmsisch-Partenkirchen, D-82467, GERMANY

FEATURES

source 1..5476

/organism="Synechococcus leopoliensis"

/strain="SAUG 1402-1"

/db_xref="taxon:32047"

/country="Germany"

gene complement(1..856)

/gene="crtQ"

CDS complement(<1..856)

/gene="crtQ"

/codon_start=1

/transl_table=11

/product="zeta carotene desaturase"

/protein_id="CA865434.1"

/db_xref="GI:6689332"

/translation="MRVAIVGAGLAGLAAALDLDVAGHQVAIYDSRPFVGGKVGSHID AGNHEMLGHLFFNYANLFAIMRKVGAPENLLPKRAHTHTFINKGGEVGEIDFEPPI GAFNGLKAFFTSQTLTKLQNALAGTSPVLRGILYEGAKMIIRALDRISFADW FRYSGSEGLSKMNPPIAYALGFIIDENISARCMITVFQMAAKTEASKLNLGASP AEFVHKPILDYIQARGATLHLRRRVEIEVTEINGTQVTGTLQADGDAVEREADYV LAACDVPGI"

988..2207

/gene="dxr"

RBS 988..992

/gene="dxr"

CDS 999..2207

/gene="dxr"

/function="Involved in isoprenoid biosynthesis"

/codon_start=1

/transl_table=11

/product="deoxyxylulose 5-phosphate reductoisomerase"

/protein_id="CA865435.1"

/db_xref="GI:6689330"

/translation="MRAVYLLGSTGSGTGTDLILQYDPRFLVGLAAGRNVALLE QIRHREPEIVADIAQLQALADLNPPLITGEGAVTEVARYGDAEIVYTGIV GCAGLPTAAIEAGKDIALANKETLIIAAGPVVPLQLQKHGVTITPADSEHSAIFQCI QGLSTHADFPDPAQVAGLRLLITASGGAFFDMPVERLSQVTVADALKHPNWSMGRKI TVDSATLMKGLVEIEAHVYLFGLDYDIDIVHPOSI IHSLEIEDTSVLAQLGWPKM RLPLLYALSWPDLTSQNSALDLVKAGSLEFFRPDHAKYPCMDLAYAAGKKGGMTPAV LNAANQVAVLFEEQIHFSDFPLRIERACDRHQTEWQQQPSLDDILAYDANARQFVQ ASYQSLESVY"

RBS 2206..2209

/note="ORF2"

CDS 2217..2816

/note="ORF2"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CA865436.1"

/db_xref="GI:6689331"

/translation="MYRYRLLFKHPYDAVCQSPDRPQOOLKDYIDVPEYVPVGR L DRUSEGLLLTNGALQHLRCHFRGHDRTYVQVREPTAALQALRQGVQIQDRT RPAKVLRDQDPQIPERDPTFRFKTPTTAWLATLQEGRNRRMTAAVGHPTLRLI RAAIATHPQPPQLQGLAPGTWRDLTAIECRLESLLR"

RBS 2827..2830

/note="ORF3"

CDS 2846..3757

/note="ORF3"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CA865437.1"

/db_xref="GI:6689332"

/translation="MTAGMSIPISEASFVAADERTAMLSLSLAQMEVAATTVGM RSQNEDFVYHYQLEHLETPSGTSVQAKGLFVLCGMMGHARGETASLEAARFKQE IFAQWTELPERSALRSVAIVNAQALFELNQAASVSGSDRMGTTLVAIILQGTVAII HVGDSRAYRSRLTGLDQTRDHEVGQLAIAQGVDPDVAIARPEARHQLTQALGPRNND WLDPEVRLEVEDTVFLCSDGLSDFDLVEEVETHTVAPLLQTRTSLQYGIQQLVEI ANEAGHDNITVAVRLRLRPLGPVI"

CDS complement(3791..4384)

/note="ORF4"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CA865438.1"

/db_xref="GI:6689333"

/translation="MLGRQDQLTADGLTQAEQLGKQWQOTCWRPQAIWTSPLQRAQ QTAIALSQGTQVPIPLLVQALTEIDGLCTLTWPEAEQRYPDLCQQLLRDRHRPPI PEATILSDCRMRAEOVCTQIITEADLWYISHGGFLQHLAALLGCDRSWGLDWPLLG AIDLSLAHEDWDTAGDRWNTSLWRWQILHWPSSAI"

RBS complement(4398..4402)

/note="ORF4"

CDS complement(4410..5476)

/note="ORF5"

/codon_start=3

/transl_table=11

/product="hypothetical protein"

/protein_id="CA865439.1"

/db_xref="GI:6689334"

/translation="ACRSLQBPFLVLSDRLSFLVLPYQSPCYFSQTPVEVLAE AYRLAQHTAHFPAQLPALARALFRFLANVAETWAFMAVLOQWPTVAASDRPS TLESDRPSRVSQADLELQAIAREVKTPLATIRTLRLLRRVLDPAQAKRLLEAI DECESDIRDFGLFFRAVELEKQESQALLARTSLTEVETVPRWKAERSLALD LQIPERPAVVSDDPTMLDQALISLVERFRNLPAGSRIBIIVSLAGAQKLRLTATEL VSDSEQPRTPASLSGLTVQTPETGSLKLEVSKNLFQALGKGLTVRRRSPQASEV ITVFLPVPSSDATTADSPR"

BASE COUNT 1135 a 1592 c 1626 g 1123 t

ORIGIN

alignment_scores:

Quality: 12.00 Length: 12

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x SLE250721 ..

Align seg 1/1 to: SLE250721 from: 1 to: 5476

176 GluGlyArgAsnArgGlnValArgArgMetThrAla 187

|||||

2631 GAAGGACCAATCGCAAGTACGGCGGATGACCGCT 2666

seq_name: gb_bal:D90899

seq documentation_block:

LOCUS D90899 133859 bp DNA BCT 07-FEB-1999

DEFINITION *Synechocystis* sp. PCC6803 complete genome, 1/27, 1-133859.

ACCESSION D90899 AB001339

VERSION D90899.1 GI:1651650

KEYWORDS

PNIL34; 2nd component required for LlaI restriction activity; ABC transporter; DNA ligase; GDP-D-mannose dehydratase; HtaR suppressor protein; N utilization substance protein; Nifs protein; UDP-N-acetyl-D-mannosaminuronic acid transferase; a negative regulator of pho regulon; adenylate kinase; anthranilate synthetase alpha-subunit; arginine decarboxylase; biopolymer transport ExbB protein; cysteine synthase; dTDP-6-deoxy-L-mannose-dehydrogenase; dihydrodipicolinate reductase; dimethyladenosine transferase; drug sensory protein A; erythrocye band 7 integral membrane protein; protein 7.2B, stomatin; ferric aerobactin receptor; ferriochrome-iron receptor; geranylgeranyl pyrophosphate synthase; glycogen (starch) synthase; guanylate kinase; hydrogenase; isoenzymes formation protein HypD; integrase-recombinase protein; iron(III) dicitrate transport system permease protein RecB; iron(III) dicitrate transport system permease protein RecC;

iron(III) dicitrate transport system permease protein FeC;
 iron(III) dicitrate transport system permease protein FeC;
 iron(III) dicitrate-binding periplasmic protein; leader peptidase
 I; malic enzyme; mutator Mutr protein; peptide methionine sulfoxide
 reductase; phosphoglycerate mutase; phosphoribosylformyl
 glycinamide synthetase II; photosystem I subunit ii; photosystem
 II 13 kD protein; photosystem II D1 protein; phycoerythrin alpha
 phycoerythrin lyase CpcF; polyribonucleotide
 nucleotidyltransferase; prenyltransferase; regulatory protein PchR;
 regulatory protein PchR; ribonuclease E; ribonuclease HII;
 thiorodoxin M; transposase; type 4 prepilin peptidase; zeaxanthin
 glucosyl transferase; PNIL34.
 Synchocystis sp. (strain: PCC6803) DNA.
 Synchocystis sp.
 Bacteria: Cyanobacteria; Chroococcales; Synchocystis.
 Tabata, S.
 Direct Submission
 Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshi
 Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure
 2; 1523-3, Yanouchino, Misarazu, Chiba 292, Japan
 (E-mail: tabata@kazusa.or.jp, Tel: +81-438-52-3933,
 Fax: +81-438-52-3934)
 2 (sites)
 Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y.,
 Miyajima, N., Hirose, M., Sugita, M., Sasamoto, S., Kimura, T.,
 Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Nariuo, K.,
 Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A.,
 Yamada, M., Yasuda, M. and Tabata, S.
 Sequence analysis of the genome of the unicellular cyanobacterium
 Synchocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions
 DNA Res. 3 (3), 109-136 (1996)
 97061201
 Potential protein coding regions were assigned on the basis of
 similarity search of the ORFs and GeneMark analysis.
 Location/Qualifiers
 1..133859
 /organism="Synchocystis sp."
 /strain="PCC6803"
 /db_xref="taxon:1143"
 1..772
 /gene="ispB"
 <1..772
 /gene="ispB"
 /note="ORF ID:slr0611"
 /codon_start=2
 /transl_table=11
 /product="prenyltransferase"
 /protein_id="BAA16579.1"
 /db_xref="GI:1651651"
 /translation="ARHRLAETEMHTASLVHDDVDEADLRNVPVNSLFDNRV
 AVLAGDELFASSWYLANLDNLEVVKLLSEVIRDFAGEILQISINRFDITDLETYLE
 KSYKATSLANSAKAGVLSADPRDVCDDHLYEYKHLGLAFQIVDDIDFTSPTEVL
 GKPSGDDISGNITAPALFAMERYPLGKIEREFAQAGLEALELVEQDGGIRRR
 ELAANQALAQHLVSLMSAPRESLLELDVYVLGRH"
 937..1494
 /note="ORF ID:slr0612"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAA16580.1"
 /db_xref="GI:1651652"
 /translation="MGRDODSEGLLLTSNGKLOHRLAHREFAHORYTFAOYEGSPT
 DEDEPLRRGTTADYTRPAIAKIITEPDPENRPPYIRYASIPISWLSITITEGRN
 ROVRMTAAVGFPLRLVRLVQIQVYTGSPQGGKKSANTWCLTLEGSLPGQWRPLTPW
 EENFCQQLLGNPNQPKKFGDRR"
 1577..2096
 /note="ORF ID:slr0613"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAA16581.1"

/db_xref="GI:1651653"
 /translation="MSYLIAVWNRRAEAEAYTTLEQAGFAQKNLTITIGTKYKTADEF
 GLVDPKQAKRAKMAIWLVPFGFAAGYCNLTIGLSTLDWAGDGNHIVGGLGAI
 GTGMSFFGVGLSGSGDSLPYRNLLQAGKYLWVYVAGGELQKQKATNLLRPLNPE
 YLQYTAPDEAFV"
 complement(2172..2873)
 /note="ORF ID:slr10558"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAA16582.1"
 /db_xref="GI:1651654"
 /translation="MSDNLTELSQLHDAEKKOLTAIAEMEGEGOGILLDYLA
 NVLEKPLAVGNVQTLNLEQETITQLQRYPTGIFLQSNQGDIDPLQBLG
 ODEFTADEITRDLCELAGASORQWLYFTEVEKFPALDLHTINALWHLHNGNFGF
 SVQRRLWASGKEFTKLWPKIKWNGSNVWTRPKGFTWDLSPAQGHLLPLNLQGRV
 AESLRHPVWSQXGW"
 complement(3192..4268)
 /gene="pNIL34"
 complement(3192..4268)
 /gene="pNIL34"
 /note="ORF ID:slr11214"
 /codon_start=1
 /transl_table=11
 /product="PNIL34"
 /protein_id="BAA16583.1"
 /db_xref="GI:1651655"
 /translation="MNTLEKPGFDEIRPGVKTPAKETILTFRYTTDDMAKMDIS
 PNEDELRAILEERVDYNNRHHFVNESFNKSWDHIDKQKLEFLEKSTAEFGF
 LLKPEGRUKNNPLLAECFNLMDSDEAHAGFNKMSDFNLSDGLFLTKSRYT
 FPKKFFIATYLSKIGYRWYITTYHLEKNPDICYPIFEFFENWQDNREFLYST
 DALMRAQPHTLNDKALMCRFFLLSVFATMYLNDTORADFYACLGARSYDKVIE
 KTNETAGRVPIILDVNNPEFYNRLETVCVSNQNLRAIDASGAPGVKALRKLPIFAS
 NGWQFKLYLKMPIAVDOLAGVR"
 complement(4451..5389)
 /note="ORF ID:slr11213"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAA16584.1"
 /db_xref="GI:1651656"
 /translation="MLSLENQRLVTGGAGFLGKQVVAQLIAGGDRPKITIPRSKDC
 DLQASACERAVENQDIVIHLAAHVGIGLNREKPAELFYDNLMMGVLIHAHQAGV
 KFEVCGTICAYKPTFPFKEEDLWNGYPEETNAPYVAKKALLVQLESYRLOYGFN
 GYLLENLYGPDNEDPDRSSHVIPALIKHYEAKAGOKQLPVWGDGSPTRFLYST
 DAARGVMGTQAYKADPVNLGTNPFISIKDLTELICELMEFEGLIWDTPNGOPR
 RCLDTTKAAGFEQAQVSLREGKNTIDWYRQRS"
 complement(5534..56511)
 /gene="rribd"
 complement(5534..5622)
 /gene="rribd"
 /note="ORF ID:slr11212"
 /codon_start=1
 /transl_table=11
 /product="GDP-D-mannose dehydratase"
 /protein_id="BAA16585.1"
 /db_xref="GI:1651657"
 /translation="MSKSKVLLTIGTGDSYLSLLEKQVGHGIIRRTSTNTD
 RIDHLYVDPHDLKALRLHYGLTDGTLRLRILEDVKPTIYNLGAQSHVRSFDSPE
 YTVDSVAMGTLLRLAIRDYQHRTGIQVREYAGSSEMFGKVEIQKETTFFPRSP
 YAKAVYGHMOTVYRESYDLFACNGILFNHESPRGFTFVTRKITRAIARIIVAGTK
 KLYLGNIDSKRWGIAKDIYRAWAWLQEQDDIVVATGETHKEYKELEIAFGVNL
 NQNVYAFDERYLRFRAVDLLIGDPAKLAQLGWEPSTVTELVHLMVEADLAVLGLT
 SPNQSGRKELMAQDMAFIRSONGHAVD"
 7229..8311
 /gene="psbA2"
 7229..8311
 /gene="psbA2"
 /note="ORF ID:slr1311"
 /codon_start=1
 /transl_table=11
 /product="photosystem II D1 protein"
 /protein_id="BAA16586.1"

gene
CDS
/db_xref="GI:1651658"
/translation="MTTILQQRSEASLNEQFCQWSTNNRIYVNGFTLMTPTLITTA
TTCFIATAPVDDIGREPAGSLNIIISGAVVPSNNAICVGHFPYLPWEARSL
DMLYNGYQVDFHFLIGCYMGROWELSYRLGMRPWICVAYSAPVSAATAVFLI
YTGOGSDTEPESLIGTGFNFMVFOAENHLMHPHMLGVAGVFGSLFSAHMSGL
VTSIVRETEGVSNGYKFGQSEETYNVAAGYFGRLIFOYASFNNSRLHFHFLG
AMPVGIWFTANGVSTMAFNANGFNQSILOSGRVIGTWADVLNRIANGFEVMEHR
NAHNPDLASGEQAPVALTAPVNG"
8492..10471
/gene="spea"
8492..10471
/gene="spea"
/note="ORF_ID:slr1312"
/codon_start=1
/transl_table=11
/product="arginine decarboxylase"
/protein_id="BAAL6587.1"
/db_xref="GI:1651659"

gene
CDS
/translation="WCEEPVPADKALGKFKKKNASWSESEALYRVEMGAPYFAI
NAAGNITVSPDGRGSLDLELVEALRORLGLPLIRFESDILADRLRLNSCFAKA
IARYNPTQAVIPVKCQQRHLVEALVRFQTSQCGLGSKPMLTALATLPPL
DROKHRTPLIICNGYKQDYLETALAKRLGHRPIITIEQLRELEWLIHQQLNKH
PMLGVARLSGSLSSSESSGDRAKLGLTMDPIVTHRELENNCLDCKMLHFK
LGTQVSDIALIEMAREASOLYVELVKAKMYRLNMGGLAVDYDGSKTYNPASKNY
NQNTANDIVAIQACELGOVSPILVSESGRAIMAHQSVLFDVLSGNOTGFSEPH
PDERNAHPLKLMCEYITITAEQIQEQLHDLQKTEASSLNFNGILSLTERGQAEQ
IHWAACKFEITROLEYIPEDQALDKIMTDIYYVNLVSFQSAPEWSLQDLFFILP
IHLNEKSPVILADLDCSDGKIDRFIDLDWVKSYLEVHPLENDGNPYILGMFLVG
AQEIMGNLNFQGOINVHVIATPQGYQIESVVRGDTMTVELGVYQVDSDDLLEGLR
RTELALNSQGITLESRLLEDYEQSLRYTYS"
complement(10622)..12631
/gene="lig"
complement(10622)..12631
/gene="lig"
/note="ORF_ID:slr1209"

alignment_scores:
Quality: 12.00 Length: 12
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x D90899 ..

Align seg 1/1 to: D90899 from: 1 to: 133859

176 GluGlyArgAsnArgGlnValArgMetThrAla 187
|||||
1231 GAGGGCGCATCGTCAGGTACGTCGATGACACGG 1266

seq_name: gb_ba2:AE001942

seq_documentation_block: 9999 bp DNA BCT 22-NOV-1999

LOCUS AE001942 Deinococcus radiodurans R1 section 79 of 229 of the complete

DEFINITION Chromosome 1.

ACCESSION AE001942 AE000513

VERSION AE001942.1 GI:6458611

KEYWORDS

SOURCE Deinococcus radiodurans.

ORGANISM Deinococcus radiodurans

REFERENCE 1 (bases 1 to 9999)

AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
Genome sequence of the radioresistant bacterium Deinococcus

radiodurans R1

Science 286 (5444), 1571-1577 (1999)

20036896

2 (bases 1 to 9999)

REFERENCE

AUTHORS

White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.

TITLE

JOURNAL

Submitted (08-NOV-1999) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

FEATURES

Location/Qualifiers

source

1..9999

/organism="Deinococcus radiodurans"

/strain="R1"

/db_xref="taxon:1299"

/chromosome="1"

complement(71..655)

/gene="DR0891"

complement(71..655)

/note="DR0891"

identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="DNA-binding response regulator"

/protein_id="AAF10469.1"

/db_xref="GI:6458612"

/translation="MRLVADDDHPLFRMGLKVALIQGDDVVAEADGFSALDADCTY

OPDALLDVKPMGTGIEVCDELRLTAPRVVSILITTPAEPALVQAPAAAGKAGYVSK

EDPESLARQLREIVAHPEVRLPOVDVPLRPRESSEVLPLLRGFSNKEIARDLGVS

PDTVHDHLARLYSKLNAGDRTEAVSRARSIGLLS"

complement(652..1761)

/gene="DR0892"

complement(652..1761)

/gene="DR0892"

/note="similar to PID:2245589 percent identity: 47.14;

identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="sensor histidine kinase"

/protein_id="AAF10470.1"

/db_xref="GI:6458613"

/translation="MTRMTTPAAPRAAAAPPERRPARLAANVAGRSRPPQLTQTOFT

LVYSLALPLNLALASGASLSWSAAAMLLVGLLSAVIGYVLSGALLRPLHRLLE

EYSGDFQFHADDPREIRSLRFAFGSLTRGCTQTRNFAFMATVHDLKPLIATG

HUTRALTEPLPDDEKRAVGQEIQTETERLLALVOOMADAHFRDVOVLHPLAGDLR

FVAEQARRLRPQAEARGLSLVSGSGQAEIDAAALERAAGVNLTEALYARHEVLT

AAPGLRVDDGGGLSADDDLAQFFNAQPAVIAAGQOYTAGTAGLGFIVRIRAEAHG

GTLYRGPPRGRPLPAAPGAPFLTFTLVLP

complement(1922..2617)

/gene="DR0893"

complement(1922..2617)

/gene="DR0893"

/note="similar to GB:U00096 PID:1787004 percent identity:

66.98; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF10471.1"

/db_xref="GI:6458614"

/translation="MVKSMQOIAMTQOKTLDOVTRFMATYSSMAAGLALTAGAVLT

AQEGSLAQVSLRLPLMLAFVLSMFQRLSAVAAGALFVGTALGTGTFSL

LFAYSPPAAVITAFVAGSAGTFLGMSVAGFVKKDLGAMGRFFLFAVLGVAMLVNFV

GSSALSLGISMIGVFLFAGLTAYDTQMLRNALSGISGQAEASINGALALYLDIFN

ITFLFLINTGNSRD"

complement(2702..4207)

/gene="DR0894"

complement(2702..4207)

/note="DR0894"

identified by Glimmer2; putative"

/codon_start=1


```

MEDLINE
REFERENCE
AUTHORS
TITLE
    91345841
    2 (bases 1 to 28206)
    Buchanan,C.E. and Ling,M.L.
    Isolation and sequence analysis of dacB, which encodes a
    sporulation-specific penicillin-binding protein in Bacillus
    subtilis
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
    92193254
    3 (bases 1 to 28206)
    Sorokin,A., Zumbstein,E., Azevedo,V., Ehrlich,S.D. and Serror,P.
    The organization of the Bacillus subtilis 168 chromosome region
    between the spoVA and spoA genetic loci, based on sequence data
JOURNAL
MEDLINE
FEATURES
    95020538
    Location/Qualifiers
    1..28206
    /organism="Bacillus subtilis"
    /strain="168"
    /sub_species="Marburg"
    /db_xref="taxon:1423"
    1..1239
    /gene="spoVAF"
    <1..1239
    /gene="spoVAF"
    /codon_start=1
    /transl_table=11
    /protein_id="AAA67472.1"
    /db_xref="GI:410115"
    /translation="VDIVENRLNNAQVEKVKLTDETQVLSGLVAVIVEGAGAFII
    DVRSYGNRPEDPEKVVARGDGFENVVNTALLRRIRDERLVRKMTKVGERSK
    TDLSCYIEDIADPLVELVEKEIASIDVGLTMADKVEEIVNQSNPPLVRYTE
    RPDVAANHVLEGRHIIIIVDTSPIIITPTLPHVQHAEEYKQAFSVGTFLRWRRFG
    ILASFLFIPNLFVLQPDLLPDNNKFTGLNKTDPHPIILQIFLADLGIETFRMAAII
    TPTALSTAMGLIAAHLIGQIAIEVLGFSPEVILYVSLAAGITFTTSPSELSATNEPS
    CPDPRCTSYKRAHRHLYSANYAMASIKSLQTPYLWLPIPNKALQVLVRTAKPG
    AKVRPSIVHPKNRLQAQPTNS"
    1158..1164
    /gene="spoVAF"
    /citation=[1]
    1245..1250
    /gene="lysa"
    /note="putative"
    1245..2664
    /gene="lysa"
    1266..1272
    /gene="lysa"
    1345..2664
    /gene="lysa"
    /codon_start=1
    /transl_table=11
    /product="diaminopimelate decarboxylase"
    /protein_id="AAA67473.1"
    /db_xref="GI:410117"
    /translation="MFLGTSRQNOGHLEIGVDALYLAEKYGTPLYVYDVALIRER
    AKSFQAFISAGLKAQVAYAKAFSVAMQIAEBEGLSLDVVSGELYTAVAGFPA
    ERHFHGNKSRLEMRMALEHRIGCIVDNFYETALLDLCKETGHSIDVLRITPGV
    EAHTHYITGGEDSKFGLDHNGQTERAIEQVLOSEHQHLLGVCHIGSQIFDTAGF
    VLAAKYFKLDEWRDSYFSVSKVNLGSGFGIRYTEDDEPLHATYEVKEIKIYAVKEN
    ASRYGFDIPEIWEIPGRSLVDAGTTLTVGSKQEPVGVQVAVYDGGMNDNRIPALY
    QAKYEAARNRIGEAHDKTVSIAGKCESSEMDLIWDIDLPEVKEGDLAVFCTGAGY
    SMANNRIRPRAVFVENGAEHLVVKRETYEDIKVLDLPFKTVGKQ"
    1610..1616
    /gene="lysa"
    /citation=[1]
    1751..1757
    /gene="lysa"
    /citation=[1]
    2496..2502
    /gene="lysa"
    /citation=[1]
    2504..2510
    /gene="lysa"

terminator
terminator
CDS
    /citation=[1]
    2666..2694
    complement(2666..2694)
    /note="ORFX19"
    complement(2693..3565)
    /note="ORFX19"
    /codon_start=1
    /transl_table=11
    /protein_id="AAA67474.1"
    /db_xref="GI:410118"
    /translation="MKRIWIGMLAAAVLLMVPKVLADAAVGDVIVITGLADLSESDK
    OKVIDEMNVPDNTATVTNKEEHEYLKYSNAOIGSRALTSSTSSITAKKSGNLVE
    THNTSGITDEMYALNMTAGVKDAKVVYTAPEVSTAAITGLIKAYEVESDEAISED
    VKQVANOELVTSSELGDKIGNERAALIAKIEEPAKNGVDPNKAIEKQVDDAASDL
    NVTLDSONQLVSLFNKMNADIDWGOVSDQLDRAKDKITKTFIESDEGKNFIQKVID
    FVSIWNAIVSIFK"
    3347..3353
    /citation=[1]
    3413..3419
    /citation=[1]
    3428..3434
    /citation=[1]
    3503..3509
    /citation=[1]
    3819..4290
    /gene="ppiB"
    3819..4250
    /gene="ppiB"
    /codon_start=1
    /product="peptidyl-prolyl isomerase"
    /protein_id="AAA67475.1"
    /db_xref="GI:410119"
    /translation="MKTGYFLLDGNKIEFELYPEAPCTVANFEKLANEGFYDGLTF
    HRVIFPGVSOQCGPHTGTGGPGTYIKCTEGNPHTEAGALSMAHAGKDTGGSOFFI
    VHEPQHLNGVHTVFGKVTSGLEFAKNMNSGDMKVEVRVG"
    3982..3988
    /gene="ppiB"
    /citation=[1]
    4261..4290
    /gene="ppiB"
    4287..4320
    /note="inverted repeat"
    5264..5467
    /note="ORFX1"
    /codon_start=1
    /transl_table=11
    /protein_id="AAA67476.1"
    /db_xref="GI:410120"
    /translation="MQVLTGLLEQYGVLFVFLMELLLALLGPGEFAGVLVYFQEKLN
    WYLSMLAAGLSCGSEHVLDVRV"
    5445..5834
    /note="ORFX2"
    /codon_start=1
    /transl_table=11
    /protein_id="AAA67477.1"
    /db_xref="GI:410121"
    /translation="MSYWIGYKLPAAIQAKAFASRGVLAVIVLFTFTVFRGQDRKI
    ELSFTFVIGGELYEEGIRLFLSHLPVHAPLESVLYPFPSEQTLTAFVIFGFCAY
    LFVRSHKTRQTVAFVFCLLSYSSSG"
    6022..6027
    /gene="sips"
    /note="putative"
    6022..7323
    /gene="sips"
    6045..6051
    /gene="sips"
    /note="putative"
    6683..6688
    /gene="sips"
    /note="putative"
    6705..6710
    /gene="sips"

conflict
-35_signal
gene
-10_signal
-35_signal
-10_signal

```

```

CDS
/note="putative"
6740..7294
/gene="slpS"
/codon_start=1
/transl_table=11
/product="signal peptidase"
/protein_id="AAA67478.1"
/db_xref="GI:410122"
/translation="MKSSENKKKSILEWAKIYAVIIVALLIRNFIFAPYVVDGDSM
YPLHNKERVVNTVXIIGFDRGDIIVLNGDDVHVYKRIIGLPGDTVMKNDQLYI
NGKVDPEYLAANKRAKQDGFHLTDGFKVYKPNKIFVVGDNRENSMDSRNLGL
FTRKQIAGTSKFEVFPNEMRKIN"
7299..7323
/gene="slpS"
7388..7393
/note="putative; ORFX4"
7410..7416
/note="putative; ORFX4"
7529..7873
/note="ORFX4"
/codon_start=1
/transl_table=11
/protein_id="AAA67479.1"
/db_xref="GI:410123"
/translation="MGRIKTKITILLVLLLAGGYMYINDIELKQVPTAIGOTLSE
EEYTIQEVKVTIDGSEYHGAENGKIIFNGKLNQDLSIDKEGDKIKAYFSKSR
IDGLIKVAKVND"
7872..7911
/note="ORFX4"
7937..7942
-10_signal
7960..7965
protein_bind
8126..8137
/note="putative"

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-252-691-7056 x BACDIA
  Align seg 1/1 to: BACDIA from: 1 to: 28206
  176 GluGlyArgAsnArgGlnValArgArgMet 185
  |||||
  17989 GAGGACCGACGACGACGACGACGACGACGATG 18018
seq_name: gb_bal:BSUB0013

seq_documentation_block:
LOCUS BSUB0013 218470 bp DNA BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 13 of 21): from 2395261
to 2613730.
ACCESSION Z99116 AL009126
VERSION Z99116.1 GI:2634723
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
REFERENCE
1 (bases 1 to 218470)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertolo, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Boursier, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
Faulstich, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A.,
Gallardon, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golligly, E.J.,
Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F.,

```

```

Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Mostl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogihara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetelle, B., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
REFERENCE
2 (bases 1 to 218470)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
FEATURES
Location/Qualifiers
1..218470
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(7..18)
/gene="cmk"
complement(7..696)
/gene="cmk"
complement(22..696)
/gene="cmk"
/EC_number="2.7.4.14"
/note="alternate gene name: jofC, ypfC"
/codon_start=1
/transl_table=11
/product="cytidylate kinase"
/protein_id="CAB14221.1"
/db_xref="GI:2634724"
/db_xref="SWISS-PROT:P38493"
/translation="MEKUSIAIDGPAAGKSTVAKIVAEKKSYYIIDTGMVRAITY
AALQNVDTDEKLAELKRDIDLIITTKDGQKVEVNGTVEALRTDEISNQVIA
AKHRSVREEMVKRQQLGKGVVMDGRDIGHVLPNAEVKIFLLASVEERAKRYEE
NVKKGPDVNYETLIEIARRDKLDSREVSPLRKABDALEIDITSLISIQEVADKILEA
VEQKSP"
complement(775..951)
/gene="ypfB"
complement(775..951)
/gene="ypfB"
/function="unknown"
/note="alternate gene name: jofB"
/codon_start=1
/transl_table=11
/protein_id="CAB14222.1"
/db_xref="GI:2634725"
/db_xref="SWISS-PROT:P38492"
/translation="MKTFFELLIKLLFIQAIIILGVQVLFHQHIEPYVSKVIOYEGV
DKMEENNRTEPKH"
complement(996..1649)
/gene="ypfA"
complement(996..1649)
/gene="ypfA"
/function="unknown"

```


only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr22>

LL22NC03-113A11 is from the human chromosome 22-specific cosmid library LL22NC03, constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550 under the auspices of the National Laboratory Gene Library Project sponsored by the US Department of Energy. The source of the flow sorted chromosomes was a human/hamster hybrid containing chromosomes Y, 22 and 9. VECTOR: lawr1216

IMPORTANT: This sequence is not the entire insert of clone LL22NC03-113A11. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone LL22NC03-113A11 is at 1 in this sequence. The true right end of clone RPL180M12 is at 100 in this sequence. The start of this sequence overlaps with sequence 282190. The end of this sequence overlaps with sequence 282248.

FEATURES

source

1. .22633
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="LL22NC03-113A11"
/clone_11b="LL22NC03"
16. .106

repeat_region

/note="MIR repeat: matches 56. .147 of consensus"
1533. .1910

repeat_region

/note="AlusX repeat: matches 1. .311 of consensus"
2252. .2565

repeat_region

/note="AlusX repeat: matches 1. .312 of consensus"
2676. .2974

repeat_region

/note="Alusq repeat: matches 1. .299 of consensus"
3015. .3179

repeat_region

/note="AluJ repeat: matches 3. .127 of consensus"
3180. .3480

repeat_region

/note="AlusX repeat: matches 1. .301 of consensus"
3481. .3590

mRNA

/note="AluJ repeat: matches 127. .243 of consensus"
join(<3961. .4021,8202. .8284,8423. .8715)

/gene="CN113A11.1"

/note="match: CDNAS: Em:AB014545; match: ESTs: Em:AA793897 Em:AA50024 Em:AA259014 Em:A1040181 Em:AA215487 Em:A1358529 Em:H55161 Em:AA976016 Em:AA215486 Em:AA632576 Em:AA927357"

/evidence="not_experimental"

/product="CN113A11.1 (KIAA0645)"

gene

3961. .9230

CDS

join(<3961. .4021,8202. .8284,8423. .8715)

/gene="CN113A11.1"

/note="match: proteins: Tr:O75140"

/codon_start=3

/evidence="not_experimental"

/product="CN113A11.1 (KIAA0645)"

/protein_id="CAB62971.1"

/db_xref="GI:6572403"

/translation="PEPEYWDRLHFOFAIAHRCGFVODKYASAFNPAENKPYOI

HTGTGVLQLPYSKRFSCQRRRSTSTNQMFCEERVGINWYNMTLTKWRSS

ATGDEKFADELLAKDFDFCINRDLVTFWTSCLKRMHASAP"

repeat_region

4628. .4671
/note="11 copies 4 mer tata 88 conserved"

repeat_region

5355. .5946
/note="296 copies 2 mer cc 56 conserved"

repeat_region

5397. .5492
/note="4 copies 24 mer 94 conserved"

repeat_region

5506. .5879
/note="17 copies 22 mer 58 conserved"

repeat_region

5576. .5851
/note="12 copies 23 mer 68 conserved"

unsure

5608. .5612
/gene="CN113A11.1"

misc_feature

5608. .5612
/note="Tandem repeat"

5608. .5612
/gene="CN113A11.1"

5608. .5612
/note="Tandem repeat; Forced join; Restriction data estimate of gap 200bp"

misc_feature

5613. .5618
/gene="CN113A11.1"

unsure

5613. .5618
/note="Tandem repeat; Single clone region"

5613. .5618
/gene="CN113A11.1"

repeat_region

5740. .5811
/note="3 copies 24 mer 94 conserved"

repeat_region

5889. .5980
/note="4 copies 23 mer 81 conserved"

misc_feature

6042. .6445
/gene="CN113A11.1"

repeat_region

6268. .6477
/note="match: GSS: Em:AAQ494872"

repeat_region

6680. .6953
/note="5 copies 42 mer 90 conserved"

repeat_region

7399. .7517
/note="AluSg1 repeat: matches 1. .270 of consensus"

repeat_region

7539. .7713
/note="AluSg/x repeat: matches 183. .301 of consensus"

repeat_region

8842. .8877
/note="AluSg/x repeat: matches 122. .296 of consensus"

repeat_region

8845. .8876
/note="18 copies 2 mer tt 80 conserved"

misc_feature

8845. .8876
/note="8 copies 4 mer ttgg 96 conserved"

polyA_signal

complement(9032..9230)

polyA_site

9211. .9216
/note="match: STS: Em:G38618"

misc_feature

9230
/gene="CN113A11.1"

repeat_region

9351. .9611
/gene="CN113A11.1"

repeat_region

9590. .9906
/note="match: GSS: Em:AAQ263906"

repeat_region

10041. .10529
/note="AluSg repeat: matches 1. .313 of consensus"

repeat_region

10630. .10939
/note="L2 repeat: matches 1922. .2482 of consensus"

repeat_region

10948. .11026
/note="AluV repeat: matches 1. .311 of consensus"

repeat_region

11070. .11363
/note="L2 repeat: matches 2622. .2710 of consensus"

repeat_region

11669. .11834
/note="AluYb8 repeat: matches 1. .308 of consensus"

repeat_region

11835. .12129
/note="AluJo repeat: matches 1. .124 of consensus"

repeat_region

12130. .12284
/note="AlusX repeat: matches 1. .295 of consensus"

repeat_region

12891. .13019
/note="AluJo repeat: matches 124. .292 of consensus"

repeat_region

13020. .13156
/note="AlusX repeat: matches 1. .135 of consensus"

repeat_region

13157. .13460
/note="AlusX repeat: matches 1. .138 of consensus"

repeat_region

13461. .13623
/note="AlusX repeat: matches 1. .303 of consensus"


```

/note="AluX repeat: matches 138..298 of consensus"
13837..14135
/note="AluX repeat: matches 1..299 of consensus"
15037..15236
/note="AluJ repeat: matches 84..302 of consensus"
15237..15526
/note="AluS repeat: matches 1..292 of consensus"
15527..15599
/note="AluJ repeat: matches 65..84 of consensus"
15600..15827
/note="AluY repeat: matches 75..300 of consensus"
15828..15887
/note="AluJ repeat: matches 1..65 of consensus"
15905..16202
/note="AluJ repeat: matches 1..305 of consensus"
16203..16303
/note="MER20 repeat: matches 101..218 of consensus"
16581..16715
/note="L2 repeat: matches 2600..2742 of consensus"
16718..17120
/note="match: GSS: Em:AQ532885"
16722..17065
/note="match: GSS: Em:AQ835686"
16870..17176
/note="AluY repeat: matches 1..311 of consensus"
17194..17504
/note="AluX repeat: matches 1..296 of consensus"
18173..18467
/note="AluX repeat: matches 1..296 of consensus"
18755..19049
/note="AluJ repeat: matches 1..303 of consensus"

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x HSN13A11
..
Align seg 1/1 to: HSN13A11 from: 1 to: 22633
98 AlaargLeuThrGlnProGlyLysArg 106
21259 GCCAGGTTAACGACGACCTGGGAAAAGG 21285

seq_name: gb_bal:SHGCP1R
seq_documentation_block:
LOCUS SHGCP1R 107379 bp DNA BCT 16-AUG-1996
DEFINITION S.hygroscopicus gene cluster for polyketide immunosuppressant rapamycin.
ACCESSION X86780
VERSION X86780.1 GI:987088
KEYWORDS ABC-transporter; antibiotic transport complexes regulator; cholesterol oxidase regulator; cystathione synthase; Cytochrome P450; helix-turn-helix protein; ketoreductase/dehydrogenase; lysine cyclodeaminase; membrane transport protein; methyltransferase; monosaccharide transporter; orfD gene; orfDD gene; orfE gene; orfF gene; orfG gene; orfH gene; orfI gene; orfJ gene; orfK gene; orfL gene; orfM gene; orfN gene; orfO gene; orfP gene; orfQ gene; orfR gene; orfS gene; orfT gene; orfU gene; orfV gene; orfW gene; orfX gene; orfY gene; orfZ gene; orfZZ gene; polyketide immunosuppressant rapamycin; polyketide synthase; pteridine-dependent dioxygenase; rapA gene; rapB gene; rapC gene; rapI gene; rapJ gene; rapK gene; rapL gene; rapM gene; rapN gene; rapO gene; rapP gene; rapQ gene; rapR gene; rapT gene; response regulator; sensory protein kinase.
SOURCE
  Streptomyces hygroscopicus.
  Streptomyces hygroscopicus
  Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
  1 (bases 1 to 107379)
  Schwecke,T., Aparicio,J.F., Molnar,I., Koenig,A., Khaw,L.E., Haydock,S.F., O'Leary,M., Caffrey,P., Cortes,J., Lester,J.B.,

```

```

Boehm,G.A., Staunton,J. and Leadlay,P.F.
The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin
Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7839-7843 (1995)
95372374
REFERENCE
  2 (bases 1 to 107379)
  Schwecke,T.
Direct Submission
Submitted (06-JUN-1995) T. Schwecke, University of Cambridge, Dept of Biochemistry, Tennis Court Road, Cambridge CB2 1QW, UK
3 (bases 1 to 107379)
Molnar,I., Aparicio,J.F., Haydock,S.F., Khaw,L.E., Schwecke,T., Koenig,A., Staunton,J. and Leadlay,P.F.
Organisation of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of genes flanking the polyketide synthase
Gene 169 (1), 1-7 (1996)
96186895
REFERENCE
  4 (bases 1 to 107379)
  Aparicio,J.F., Molnar,I., Schwecke,T., Koenig,A., Haydock,S.F., Khaw,L.E., Staunton,J. and Leadlay,P.F.
Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase
Gene 169 (1), 9-16 (1996)
96186896
REFERENCE
  5 (bases 1 to 107379)
  Haydock,S.F., Aparicio,J.F., Molnar,I., Schwecke,T., Koenig,A., Marsden,A.F.A., Galloway,I.S., Staunton,J., Leadlay,P.F., Boehm,G.A., Staunton,J. and Leadlay,P.F.
Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases
Unpublished
  Location/Qualifiers
    1..107379
      /organism="Streptomyces hygroscopicus"
      /strain="NRRL 5491"
      /db_xref="taxon:1912"
    1..107379
      /note="biosynthetic gene cluster"
      /product="polyketide immunosuppressant rapamycin"
      complement(1..871)
      /gene="orfzz"
      /codon_start=1
      /transl_table=11
      /product="monosaccharide transporter"
      /protein_id="CAA60449.1"
      /db_xref="GI:987089"
      /db_xref="SPTREMBL:Q54286"
      /translation="MAPPSVRSVADVVALIDGLGKITGRAQIINFVFGGLFLDAYS
      NAAAGLGPMTOMELTSTQVSLTATAPALAILFNLGGWLATIGRPVPLLIARL
      FAIAGALLAFAAGDTFVWGLRVLVGAYGMDFAVAMALLAETPAKLGRLNMQAV
      WYIATSNLVALVFNLDVGADINRWSVGSAAVAVALLIGOWMLKESPTWLAKG
      RLGEATNLDKIYGIKAVAGTPEMTRTPAEPVIGLROAGRLFRGEYLPRTLLSVI
      SUGSQMQYFVAGWY"
      complement(1..871)
      /gene="orfzz"
      complement(1150..2319)
      /gene="orfz"
      complement(1150..2319)
      /gene="orfz"
      /codon_start=1
      /transl_table=11
      /protein_id="CAA60450.1"
      /db_xref="GI:987090"
      /db_xref="SPTREMBL:Q54287"
      /translation="MKFLYDDESFSFEALRAAGYAAAGGADLGEVLVTCRQIPGDEE
      ANSAQWATAARIETGRDLAAGHRSAREALLRASNRYRTADFRRNPADAESA
      RLAKASQOFTFADAAPLLDTPARALRIPYEDTLPGLFLVDDSGTPTPTVLFHGGYDS
      TLEEDYLALACALRGYNYIAFDGPGQGTVREOGLHFPPDWEAVVTVPVDFALPLP
      EYDAQQLVLMGSLGGLAARAAAEHRYVAACVLDGVYDFHVAIATAGRAASTGG
      MEALMAQNTWVRWVNRGRTWTFGSCVDLVAARAEPTWAGIADTRITCPTLVLEAEND

```


Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wymann,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-Oct-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

TITLE

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 140: contig of 140 bp in length
* gap of unknown length
* 141 4575: contig of 4435 bp in length
* gap of unknown length
* 4576 11055: contig of 6480 bp in length
* gap of unknown length
* 11056 19208: contig of 8153 bp in length
* gap of unknown length
* 19209 27069: contig of 7861 bp in length
* gap of unknown length
* 27070 35519: contig of 8450 bp in length
* gap of unknown length
* 35520 41379: contig of 5860 bp in length
* gap of unknown length
* 41380 49140: contig of 7761 bp in length
* gap of unknown length
* 49141 58881: contig of 9741 bp in length
* gap of unknown length
* 58882 69901: contig of 11020 bp in length
* gap of unknown length
* 69902 81612: contig of 11711 bp in length
* gap of unknown length
* 81613 95329: contig of 13717 bp in length
* gap of unknown length
* 95330 124361: contig of 29032 bp in length
* gap of unknown length
* 124362 158329: contig of 33968 bp in length.
Location/Qualifiers

FEATURES

Source

1. .158329
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2_L_1"
/clone_lib="RPCI-11 Human Male BAC"
BASE COUNT 47572 a 32248 c 30585 g 47923 t 1 others
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AC011098/rev ..

Align seg 1/1 to reverse of: AC011098 from: 1 to: 158329

98 AlaArgLeuThrGlnProGlyLysArg 106

97207 GCTAGACTACACAGCCCTGGGAAAGA 97181

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

OM of: US-09-252-691-7056 to: N_Geneseq_36.* out_format : pfs

Date: Jun 10, 2000 4:02 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame_pzn.model -DEV-xlp
-Q/cgn2_1/USPTO-SPOT/US09252691/runat_05062000_101738_1291/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=oligo.rng -GAPOP=4.500
-GAPOP=4.500 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPOP=0.050 -XGAPOP=60.000 -XGAPOP=60.000
-GAPOP=6.000 -FGAPOP=7.000 -YGAPOP=60.000 -YGAPOP=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:

Query: US-09-252-691-7056

Query length: 222

Database: N_Geneseq_36.*

Database sequences: 311585

Database length: 125096042

Search time (sec): 60.370000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
N_Geneseq_36:V74878	-	10.00	163.81	0.1595	676	! Staphylococcus aureus contig SE
N_Geneseq_36:X20509	+	8.00	105.72	274.42	12838	! Polynucleotide sequence from
N_Geneseq_36:X069432	+	7.00	128.49	14.79	46	! Human heat shock protein (hsp70)
N_Geneseq_36:X63984	+	7.00	128.49	14.79	46	! Human hsp70B gene TFIIID binding
N_Geneseq_36:X17182	+	7.00	128.49	14.79	46	! Test sequence from human hsp70B
N_Geneseq_36:X21143	+	7.00	115.16	81.74	286	! Human hsp70B promoter. Spatial
N_Geneseq_36:X24918	+	7.00	111.91	124.13	447	! Human lymphoma cell line U937 c
N_Geneseq_36:X15049	+	7.00	109.60	166.79	613	! Rhinopneumonitis appendiculatus pr
N_Geneseq_36:X21289	+	7.00	109.47	169.59	624	! H. pylori cytoplasmic protein C
N_Geneseq_36:X21163	+	7.00	109.28	173.91	641	! H. pylori ORF 09gell21oorfi. He
N_Geneseq_36:X15049	+	7.00	109.13	177.21	654	! DNA encoding a GST-MIG fusion pr
N_Geneseq_36:X15057	+	7.00	109.12	177.21	654	! DNA encoding GST-MIG fusion pr
N_Geneseq_36:X83108	+	7.00	109.12	177.46	655	! Human C-type lectin-1 (CTL-1) c
N_Geneseq_36:X86546	+	7.00	108.99	180.50	667	! Wilson disease exon 5/intron b
N_Geneseq_36:X49570	+	7.00	108.67	188.09	697	! Human lymphoma cell line U937 c
N_Geneseq_36:X54641	+	7.00	108.67	188.09	697	! Nucleotide sequence encoding a
N_Geneseq_36:X06949	+	7.00	107.79	210.47	786	! Bacillus licheniformis xylogluc
N_Geneseq_36:X69444	+	7.00	107.35	219.72	823	! Banana fruit ripening-related c
N_Geneseq_36:X04554	+	7.00	107.37	221.97	832	! DNA encoding a human MGL18 pr
N_Geneseq_36:X54646	+	7.00	107.06	231.19	869	! Human IL-6 receptor antisense c
N_Geneseq_36:X54649	+	7.00	107.06	231.19	869	! Human IL-6 receptor antisense c
N_Geneseq_36:X60096	+	7.00	105.76	273.01	1038	! Serpin proteinase inhibitor-2
N_Geneseq_36:X73345	+	7.00	105.61	278.18	1059	! DNA encoding lipase modulator
N_Geneseq_36:X74709	+	7.00	105.07	298.28	1141	! Staphylococcus aureus contig s
N_Geneseq_36:X52017	+	7.00	104.82	308.05	1181	! Helicobacter polypeptide GHPO
N_Geneseq_36:X25104	+	7.00	104.72	311.95	1197	! H. pylori cytoplasmic protein
N_Geneseq_36:X34569	+	7.00	104.64	315.37	1211	! H. vaccae antigen 85A encoding
N_Geneseq_36:X05610	+	7.00	104.23	332.15	1280	! nlfH gene and promoter region
N_Geneseq_36:X50059	+	7.00	104.23	332.15	1281	! Sequence of the Fe-protein sub
N_Geneseq_36:X050979	+	7.00	104.00	342.33	1322	! CRABP-II gene 5'-end. Human c
N_Geneseq_36:X82138	+	7.00	103.47	366.26	1421	! Gene encoding tropoelastin. Ch
N_Geneseq_36:X15846	+	7.00	103.24	377.34	1467	! Nucleic acid encoding CrmA. Ch
N_Geneseq_36:X33568	+	7.00	103.23	377.58	1468	! Cowpox virus CrmA gene. New nu
N_Geneseq_36:X72795	+	7.00	102.46	416.67	1631	! Beta-1 integrin modulator cDNA
N_Geneseq_36:X62311	+	7.00	102.39	420.74	1648	! Rat amelotin-2 gene. Enamel matr
N_Geneseq_36:X29052	+	7.00	101.99	442.67	1740	! Alcohol and/or aldehyde dehydro
N_Geneseq_36:X22824	+	7.00	101.82	452.42	1781	! Sequence encoding ovine interl
N_Geneseq_36:X32292	+	7.00	101.64	462.87	1825	! M. grisea_PTH2 gene coding seq

N_Geneseq_36:T62310	-	7.00	101.20	489.87	1939	Rat amelotin-1 gene. Enamel m
N_Geneseq_36:T73350	+	7.00	100.87	510.88	2028	Lipase gene expression cons
N_Geneseq_36:V61572	-	7.00	100.72	521.24	2104	Human Kv potassium channel
N_Geneseq_36:V61571	-	7.00	100.61	528.77	2104	Human Kv potassium channel
N_Geneseq_36:V61573	-	7.00	100.61	528.77	2104	Human Kv potassium channel
N_Geneseq_36:Q67062	-	7.00	100.50	536.05	2135	Rabphilin-3A cDNA. Target p
N_Geneseq_36:Q48170	-	7.00	100.26	552.71	2206	Sequence encoding a 90K tum
N_Geneseq_36:Q48888	-	7.00	100.26	552.71	2206	Sequence encoding an immuno
N_Geneseq_36:Q48161	-	7.00	100.26	552.71	2206	DNA encoding tumour associa
N_Geneseq_36:Q22147	-	7.00	100.21	556.46	2222	RUBPCase gene. RUBPCase gene
N_Geneseq_36:T81514	+	7.00	100.20	557.16	2225	Human Immune Cell Cytokine-
N_Geneseq_36:T02367	+	7.00	99.08	643.18	2594	Meloni oncotrans cDNA clone
N_Geneseq_36:T65421	+	7.00	99.21	718.97	2922	Human p100 protein coding s
N_Geneseq_36:X03030	+	7.00	98.18	721.73	2934	Human IL-1ra BAC contiguous
N_Geneseq_36:X32291	+	7.00	98.10	729.09	2966	M. grisea_PTH2 gene sequenc
N_Geneseq_36:Q52536	+	7.00	98.04	735.30	2993	Encodes Subtilisin-like ser
N_Geneseq_36:V56991	-	7.00	97.28	810.21	3320	Maize starch phosphorylase
N_Geneseq_36:V36957	-	7.00	96.98	842.13	3460	Nucleotide sequence of huma
N_Geneseq_36:V57515	+	7.00	96.86	854.87	3516	Zcyto7 cytokine receptor e
N_Geneseq_36:X32295	+	7.00	96.23	927.23	3835	M. grisea_PTH11 gene sequen
N_Geneseq_36:X19491	+	7.00	95.33	1.0e+03	4337	Human secreted protein clon
N_Geneseq_36:V04633	+	7.00	94.76	1.1e+03	4692	Porcine phosphoinositide 3O
N_Geneseq_36:V74099	+	7.00	94.76	1.1e+03	4692	Porcine G-protein regulated
N_Geneseq_36:Q29703	-	7.00	94.11	1.2e+03	5125	IRS-1. Purified nucleic aci
N_Geneseq_36:T68952	-	7.00	94.08	1.2e+03	5150	Blackcurrant fruit-specific
N_Geneseq_36:V08418	-	7.00	93.06	1.4e+03	5924	RCMV strain AD169 genomic f
N_Geneseq_36:Q75977	-	7.00	92.78	1.4e+03	6151	Insulin receptor substrate-
N_Geneseq_36:X55274	+	7.00	91.80	1.6e+03	7036	Human factor-related antisense
N_Geneseq_36:X33180	+	7.00	91.05	1.8e+03	7797	Cowpox virus bsr full length
N_Geneseq_36:X55272	+	7.00	91.05	1.8e+03	7797	Human receptor-related anti
N_Geneseq_36:V33476	-	7.00	88.09	2.6e+03	11703	Sindbis virus clone TR399
N_Geneseq_36:X20501	+	7.00	86.66	3.2e+03	14244	Polynucleotide sequence fr
N_Geneseq_36:Q40480	+	7.00	85.86	3.5e+03	15894	Attenuated measles virus s
N_Geneseq_36:V18264	+	7.00	85.86	3.5e+03	15894	Measles virus Edmonston va
N_Geneseq_36:V18269	+	7.00	85.86	3.5e+03	15894	Measles virus Rubenox va
N_Geneseq_36:V18270	+	7.00	85.86	3.5e+03	15894	Measles virus Moraten vacci
N_Geneseq_36:V18271	+	7.00	85.86	3.5e+03	15894	Measles virus Zagreb vacci
N_Geneseq_36:V18272	+	7.00	85.86	3.5e+03	15894	Measles virus Aik-C vaccini
N_Geneseq_36:V18265	+	7.00	85.86	3.5e+03	15894	Measles virus 1977 isolate
N_Geneseq_36:Q86154	+	7.00	85.51	3.7e+03	16556	Eukaryotic layered vector
N_Geneseq_36:X30787	+	7.00	85.51	3.7e+03	16556	Alphavirus-based eukaryoti
N_Geneseq_36:V42364	+	7.00	85.51	3.7e+03	16556	Representative eukaryotic
N_Geneseq_36:V60125	-	7.00	85.51	3.7e+03	16556	Representative eukaryotic
N_Geneseq_36:V70684	-	7.00	85.51	3.7e+03	16556	Representative eukaryotic
N_Geneseq_36:T08126	+	7.00	82.67	5.3e+03	24699	Polynucleotide sequence fr
N_Geneseq_36:Q67902	-	7.00	82.07	5.7e+03	26598	Syndecan gene. Syndecan st
N_Geneseq_36:V15946	-	7.00	82.07	5.7e+03	26700	Mouse syndecan-1 gene sequen
N_Geneseq_36:V81283	-	7.00	82.07	5.7e+03	26700	Mouse syndecan-1 encoding
N_Geneseq_36:X23520	+	7.00	78.18	9.3e+03	45346	Human kidney aminopeptidas
N_Geneseq_36:T58840_4	+	7.00	71.74	2.1e+04	110000	Continuation (5 of 6) of
N_Geneseq_36:V30458_3	+	7.00	71.74	2.1e+04	110000	Continuation (4 of 6) of
N_Geneseq_36:X53491	+	7.00	71.74	2.1e+04	110000	Continuation (4 of 6) of
N_Geneseq_36:T78260	-	7.00	71.42	2.2e+04	114955	Human adenosine A1 recept
N_Geneseq_36:X54050	-	6.00	116.64	67.66	19	Human IL6 receptor antisense
N_Geneseq_36:X47307	+	6.00	116.26	70.99	20	PCR primer for human globin s
N_Geneseq_36:T47900	+	6.00	116.26	70.99	20	Human globin gene primer RS-4
N_Geneseq_36:V31644	+	6.00	115.57	77.61	22	Nucleotide sequence of the CT
N_Geneseq_36:V34649	-	6.00	114.63	87.47	25	Cylindrocyclodextrin flori
N_Geneseq_36:X19595	+	6.00	114.07	94.00	27	Human genomic DNA target sequ
N_Geneseq_36:X19598	+	6.00	114.07	94.00	27	Human genomic DNA target sequ
N_Geneseq_36:V91548	-	6.00	113.30	103.74	30	Human C-rat hammerhead riboz
N_Geneseq_36:V91548	-	6.00	113.30	103.74	30	PCR primer IVS-1 nt110-N to d
N_Geneseq_36:Q48866	-	6.00	113.30	103.74	30	PCR primer IVS-1 nt110-M to d
N_Geneseq_36:T38358	-	6.00	113.30	103.74	30	Neutral protease npr gene pri
N_Geneseq_36:X19606	+	6.00	113.30	103.74	30	Human genomic DNA target sequ
N_Geneseq_36:Q66689	+	6.00	113.07	106.97	31	Human genomic DNA polymorphic
N_Geneseq_36:Q66689	+	6.00	112.83	110.20	32	Primer C2 to amplify Sali-Stu
N_Geneseq_36:X19332	-	6.00	112.61	113.42	33	Human Fc-epsilon-R1-alpha PCR
N_Geneseq_36:Q97741	+	6.00	111.97	123.04	36	Primer P283 to subclone T-gon
N_Geneseq_36:T52591	+	6.00	111.97	123.04	36	Human ICAM hammerhead ribozym
N_Geneseq_36:Q42536	+	6.00	111.78	126.23	37	Epidermin structural gene epi

N_Geneseq_36:T76275	-	6.00	111.39	132.60	39	Human neutrophil elastase antise	N_Geneseq_36:T11816	+	6.00	97.17	821.67	274	M. kansasii ATCC 22478 (-ITG
N_Geneseq_36:X54071	-	6.00	111.39	132.60	39	Neutrophil elastase (Medullasin)	N_Geneseq_36:V90164	+	6.00	97.12	827.28	276	EST clone DF478. New polynuc
N_Geneseq_36:Q69433	+	6.00	110.19	154.75	46	Human heat shock protein (hsp70B)	N_Geneseq_36:V05724	+	6.00	97.01	838.49	280	Nucleotide sequence of the 3
N_Geneseq_36:T63895	+	6.00	110.19	154.75	46	Human hsp70B gene TFIIID binding	N_Geneseq_36:X39526	+	6.00	96.99	841.29	281	Human secreted protein 5' ES
N_Geneseq_36:X17183	+	6.00	109.88	161.03	48	Test sequence from human hsp70B	N_Geneseq_36:Q92004	+	6.00	96.91	849.69	284	Synthetic gene encoding the
N_Geneseq_36:Q50128	+	6.00	109.88	161.03	48	Ruthenium-labelled, elongated ba	N_Geneseq_36:Q10102	+	6.00	96.88	852.49	285	SP-IT gene. Functional, inse
N_Geneseq_36:Q89014	-	6.00	109.88	161.03	48	VSGF 2'-NH2-RNA nucleic acid lig	N_Geneseq_36:Q26399	+	6.00	96.83	858.09	287	PCIB4223. Insecticidal bacul
N_Geneseq_36:T74685	+	6.00	109.43	170.43	51	SAC69 used in construction of GC	N_Geneseq_36:X40575	+	6.00	96.44	902.76	303	Human secreted protein 5' ES
N_Geneseq_36:T74687	+	6.00	109.43	170.43	51	SAC75 used in construction of GC	N_Geneseq_36:V87842	+	6.00	96.41	905.55	304	EST clone EY215. New polynuc
N_Geneseq_36:Q23186	+	6.00	109.29	173.56	52	HPV probe 39. Compsns. for ampli	N_Geneseq_36:T19157	+	6.00	96.36	911.12	306	Human gene signature HUMG500
N_Geneseq_36:T29881	+	6.00	109.15	176.68	53	Human papillomavirus detection p	N_Geneseq_36:V15619	+	6.00	96.34	913.91	307	Tomato partial invertase inh
N_Geneseq_36:T76532	+	6.00	109.02	179.79	54	Staphylococcus aureus contig SEQ	N_Geneseq_36:V87619	+	6.00	96.34	913.91	307	EST clone EC32. New polynuc
N_Geneseq_36:V41849	+	6.00	108.75	186.02	56	Nucleotide sequence of ribozyme	N_Geneseq_36:Q60364	+	6.00	96.18	933.39	314	Human brain Expressed Sequen
N_Geneseq_36:T17695	+	6.00	108.62	189.12	57	RNA-RNA recombination construct	N_Geneseq_36:Q32885	-	6.00	96.15	936.17	315	Korean hepatitis C virus cDN
N_Geneseq_36:Q22524	+	6.00	107.23	226.14	69	pAD-CMV1 primer EBI-2585. O-glyc	N_Geneseq_36:T76274	+	6.00	96.11	941.73	317	Human neutrophil elastase (m
N_Geneseq_36:Q20741	+	6.00	107.23	226.14	69	pAD-CMV1 primer EBI-2585. O-glyc	N_Geneseq_36:X54665	+	6.00	96.11	941.73	317	Neutrophil elastase (Medulla
N_Geneseq_36:T11271	+	6.00	107.23	226.14	69	Hepatitis C virus partial 5'-UTR	N_Geneseq_36:V69002	+	6.00	96.08	944.51	318	DNA molecule encoding a poly
N_Geneseq_36:T71502	+	6.00	107.02	232.27	71	Glioblastoma U251 cell line lig	N_Geneseq_36:V86945	+	6.00	95.99	955.62	322	EST clone AY398. New polynuc
N_Geneseq_36:T59114	+	6.00	106.82	238.38	73	DNA polymerase ligand to a therm	N_Geneseq_36:V90442	+	6.00	95.97	958.40	323	Nucleotide sequence of clone
N_Geneseq_36:V11323	+	6.00	105.37	286.94	89	pDR-08 D-Arginine binding RNA se	N_Geneseq_36:Q39306	+	6.00	95.88	969.50	327	Expressed sequence tag human
N_Geneseq_36:V11321	+	6.00	105.29	289.96	90	pDR-87 D-Arginine binding RNA se	N_Geneseq_36:Q59318	-	6.00	95.88	969.50	327	Human brain Expressed Sequen
N_Geneseq_36:V11325	+	6.00	105.21	292.97	91	pDR-17 D-Arginine binding RNA se	N_Geneseq_36:T70856	-	6.00	95.84	975.04	329	Staphylococcus aureus contig
N_Geneseq_36:X17693	+	6.00	105.13	295.98	92	Recombinant BMV RNA3 of PN10(-)	N_Geneseq_36:V70856	-	6.00	95.81	977.81	330	Internal transcribed spacer
N_Geneseq_36:Q84567	+	6.00	104.52	320.00	100	HSV type 1 IE4 and IE4 pre-mess	N_Geneseq_36:V11616	-	6.00	95.68	994.44	336	Homo sapiens adult brain clo
N_Geneseq_36:T48787	+	6.00	104.17	334.94	105	Polyclonal anti-ferritin binder	N_Geneseq_36:T20129	+	6.00	95.64	999.98	338	Probe (30) for microbial gen
N_Geneseq_36:T76283	+	6.00	104.10	337.92	106	Staphylococcus aureus contig SE	N_Geneseq_36:Q78119	+	6.00	95.60	1.0e+03	338	Hepatitis C virus NS5B regio
N_Geneseq_36:T39645	+	6.00	103.63	358.76	113	Oligonucleotide used to constru	N_Geneseq_36:Q78120	+	6.00	95.60	1.0e+03	340	Hepatitis C virus NS5B regio
N_Geneseq_36:T39636	+	6.00	103.63	358.76	113	Oligonucleotide used to constru	N_Geneseq_36:T26389	+	6.00	95.28	1.0e+03	355	Human gene signature HUMG508
N_Geneseq_36:X03882	+	6.00	103.63	358.76	113	Human papillomavirus 11 L1 open	N_Geneseq_36:V24905	+	6.00	95.24	1.1e+03	357	H. pylori secreted protein O
N_Geneseq_36:X03891	+	6.00	103.63	358.76	113	Human papillomavirus 11 L1 open	N_Geneseq_36:T28900	+	6.00	95.22	1.1e+03	358	Human gene signature HUMG504
N_Geneseq_36:T48587	+	6.00	103.44	367.66	116	Catalytic DNA, seq09. Nucleic a	N_Geneseq_36:T24847	+	6.00	95.06	1.1e+03	366	H. pylori ORF 04gpl123orf11
N_Geneseq_36:T48598	+	6.00	103.44	367.66	116	Catalytic DNA, seq10. Nucleic a	N_Geneseq_36:T15397	+	6.00	94.73	1.1e+03	383	Papillomavirus major capsid
N_Geneseq_36:T48599	+	6.00	103.44	367.66	116	Catalytic DNA, seq10. Nucleic a	N_Geneseq_36:T84381	+	6.00	94.67	1.1e+03	386	Friedreich's ataxia STM7 gen
N_Geneseq_36:V13502	+	6.00	103.31	373.59	118	Hepatitis C virus nucleic acid	N_Geneseq_36:T15598	+	6.00	94.67	1.1e+03	386	Papillomavirus major capsid
N_Geneseq_36:T84383	+	6.00	103.13	382.47	121	Friedreich's ataxia STM7 gene e	N_Geneseq_36:V15598	+	6.00	94.57	1.1e+03	386	Papillomavirus major capsid
N_Geneseq_36:T73909	-	6.00	102.55	411.97	131	Synthetic oligonucleotide encod	N_Geneseq_36:T19086	+	6.00	94.54	1.2e+03	393	Human gene signature HUMG500
N_Geneseq_36:T88358	+	6.00	102.55	411.97	131	Synthetic oligonucleotide encod	N_Geneseq_36:T19081	+	6.00	94.48	1.2e+03	396	Human gene signature HUMG500
N_Geneseq_36:T48436	+	6.00	102.23	429.59	137	Human gene signature HUMG50932	N_Geneseq_36:Q46001	+	6.00	94.45	1.2e+03	398	NPS-331. Insecticidal peptid
N_Geneseq_36:T11267	+	6.00	102.17	432.53	138	Hepatitis C virus partial 5'-UTR	N_Geneseq_36:Q59672	-	6.00	94.45	1.2e+03	398	Human brain Expressed Sequen
N_Geneseq_36:T11269	+	6.00	102.07	438.39	140	Hepatitis C virus partial 5'-UTR	N_Geneseq_36:Q59673	-	6.00	94.45	1.2e+03	398	Polyptide component of spi
N_Geneseq_36:Q77578	+	6.00	101.86	450.10	144	Human genome fragment. (Pfeffe	N_Geneseq_36:T01514	+	6.00	94.45	1.2e+03	398	Hepatitis C virus plasmid DN
N_Geneseq_36:N92934	+	6.00	101.52	470.54	151	Sequence of palindromic region	N_Geneseq_36:T01514	+	6.00	94.45	1.2e+03	398	Hepatitis C virus plasmid DN
N_Geneseq_36:V25921	+	6.00	101.28	485.10	156	Tegenaria agrestis NPS-331 codi	N_Geneseq_36:T47263	+	6.00	94.45	1.2e+03	398	(DSM 10103) human papillomav
N_Geneseq_36:X17679	+	6.00	100.83	514.13	166	Recombinant BMV RNA3 of PN8(-)	N_Geneseq_36:V25920	+	6.00	94.45	1.2e+03	398	Tegenaria agrestis NPS-331 c
N_Geneseq_36:V88880	+	6.00	100.87	531.50	172	EST clone H2109. New polynucleo	N_Geneseq_36:T78209	+	6.00	94.41	1.2e+03	400	Staphylococcus aureus contig
N_Geneseq_36:Q79436	-	6.00	100.36	545.94	177	HCV isolate GB116 5' UTR (-244	N_Geneseq_36:T78158	+	6.00	94.41	1.2e+03	400	Staphylococcus aureus contig
N_Geneseq_36:Q68066	-	6.00	100.36	545.94	177	HCV isolate GB116 5' UTR (-244	N_Geneseq_36:T77951	+	6.00	94.41	1.2e+03	400	Staphylococcus aureus contig
N_Geneseq_36:Q68067	-	6.00	100.36	545.94	177	HCV isolate BE97 5' UTR (-244 t	N_Geneseq_36:X40444	+	6.00	94.30	1.2e+03	406	Human secreted protein 5' ES
N_Geneseq_36:Q68068	-	6.00	100.36	545.94	177	HCV isolate BE95 5' UTR (-244 t	N_Geneseq_36:Q59351	+	6.00	94.28	1.2e+03	407	Human brain Expressed Sequen
N_Geneseq_36:Q68069	+	6.00	100.36	545.94	177	HCV isolate BE95 5' UTR (-244 t	N_Geneseq_36:T47356	+	6.00	94.23	1.2e+03	410	(DSM 10097) human papillomav
N_Geneseq_36:Q98249	+	6.00	100.23	554.59	180	Exon 9 of rice pyruvate orthoph	N_Geneseq_36:X20912	-	6.00	94.14	1.2e+03	415	Polynucleotide sequence from
N_Geneseq_36:Q43067	+	6.00	100.07	566.11	184	-255 to -62 region of 5' non-c	N_Geneseq_36:T30730	-	6.00	94.07	1.2e+03	419	Mouse cryptidin 5 gene. New I
N_Geneseq_36:Q43068	+	6.00	100.07	566.11	184	-255 to -62 region of 5' non-c	N_Geneseq_36:V57210	+	6.00	93.98	1.2e+03	424	Clone R1.14 from mutant mult
N_Geneseq_36:T77510	+	6.00	100.07	566.11	184	Staphylococcus aureus contig SE	N_Geneseq_36:V89078	+	6.00	93.95	1.2e+03	426	EST clone CB15. New polynuc
N_Geneseq_36:T11272	+	6.00	100.00	571.87	186	Hepatitis C virus partial 5'-UTR	N_Geneseq_36:Q84934	+	6.00	93.92	1.2e+03	428	Human gata-globin gene. Line
N_Geneseq_36:X33647	+	6.00	99.69	594.85	194	DNA tandem-nucleotide repeat l	N_Geneseq_36:X40584	+	6.00	93.83	1.3e+03	433	Human secreted protein 5' ES
N_Geneseq_36:X11504	+	6.00	99.32	623.49	204	Human biallelic polymorphic DNA	N_Geneseq_36:X51917	+	6.00	93.78	1.3e+03	436	Human secreted protein 5' ES
N_Geneseq_36:X11011	+	6.00	98.61	683.34	225	Sequence encoding AAT. Functio	N_Geneseq_36:T60776	+	6.00	93.54	1.3e+03	449	Galectin 10 and 10SV related
N_Geneseq_36:Q05051	+	6.00	98.57	686.19	226	Recombinant AAT gene for insed	N_Geneseq_36:Q70316	+	6.00	93.47	1.3e+03	457	Oxidoreducing avermectin DNA
N_Geneseq_36:T89401	+	6.00	98.54	689.03	227	Methods for diagnosing Friedre	N_Geneseq_36:Q70316	+	6.00	93.44	1.3e+03	457	Oxidoreducing avermectin DNA
N_Geneseq_36:X40618	+	6.00	98.26	714.55	236	Human secreted protein 5' EST S	N_Geneseq_36:Q63396	+	6.00	93.41	1.3e+03	459	Consensus sequence of mannur
N_Geneseq_36:T74055	+	6.00	98.17	723.05	239	Partial sequence of HCV subtype	N_Geneseq_36:T65074	+	6.00	93.39	1.3e+03	460	Canine genomic microsatellit
N_Geneseq_36:V99109	+	6.00	98.08	731.53	242	DNA methyltransferase genomic f	N_Geneseq_36:X17811	+	6.00	93.39	1.3e+03	460	Microsatellite repeat sequen
N_Geneseq_36:X15319	+	6.00	98.05	734.36	243	DNA encoding an enzyme having a	N_Geneseq_36:T60776	+	6.00	93.26	1.4e+03	468	Mabinlin MBLII cDNA from Cap
N_Geneseq_36:Q60591	+	6.00	97.96	742.84	246	Human brain Expressed Sequence	N_Geneseq_36:T19047	+	6.00	93.20	1.4e+03	472	Human gene signature HUMG500
N_Geneseq_36:V86817	+	6.00	97.81	751.31	249	EST clone BG386. New polynucleo	N_Geneseq_36:T19048	-	6.00	93.20	1.4e+03	472	Human gene signature HUMG500
N_Geneseq_36:X12503	+	6.00	97.81	756.96	251	Human biallelic polymorphic DNA	N_Geneseq_36:X20812	-	6.00	93.14	1.4e+03	476	Polynucleotide sequence from
N_Geneseq_36:X12504	+	6.00	97.81	756.96	251	Human biallelic polymorphic DNA	N_Geneseq_36:Q67467	+	6.00	93.09	1.4e+03	479	CDNA fragment used in the sy
N_Geneseq_36:X10304	+	6.00	97.81	756.96	251	Human biallelic polymorphic DNA	N_Geneseq_36:V88503	+	6.00	93.06	1.4e+03	481	EST clone EK341. New polynuc
N_Geneseq_36:X10305	+	6.00	97.81	756.96	251	Human biallelic polymorphic DNA	N_Geneseq_36:Q32446	-	6.00	93.03	1.4e+03	483	HCV core-envelope clone S1-2
N_Geneseq_36:V02955	+	6.00	97.61	776.69	258	Hepatitis GB virus (HGV) type C	N_Geneseq_36:Q32447	-	6.00	93.03	1.4e+03	483	HCV core-envelope clone S1-2
N_Geneseq_36:Q92981	+	6.00	97.44	793.58	264	M2 subunit of human ribonucleo	N_Geneseq_36:Q32453	-	6.00	93.03	1.4e+03	483	HCV core-envelope clone S1-1
N_Geneseq_36:V05736	+	6.00	97.41	796.39	265	Nucleotide sequence of the 5' p	N_Geneseq_36:X54661	-	6.00	92.94	1.4e+03	489	Neutrophil elastase (Medulla

N_Geneseq_36:V89284	6.00	92.93	1.4e+03	490	! EST clone CG279. New polynucleot	N_Geneseq_36:Q98296	+	6.00	89.32	2.2e+03	804	! Recombinant abrin A variant
N_Geneseq_36:V87761	6.00	92.83	1.4e+03	490	! EST clone EQ218. New polynucleot	N_Geneseq_36:T10398	+	6.00	89.29	2.3e+03	807	! Recombinant abrin A variant
N_Geneseq_36:V44834	6.00	92.86	1.4e+03	495	! EST clone HMOAC87R related to FCR-1	N_Geneseq_36:N80722	+	6.00	89.27	2.3e+03	809	! Canine SP5 cDNA Clone, pD6K
N_Geneseq_36:N95402	6.00	92.80	1.4e+03	499	! Sequence of new sigma-mu switch	N_Geneseq_36:N70716	+	6.00	89.27	2.3e+03	809	! Sequence of a canine SP-5 cl
N_Geneseq_36:N83787	6.00	92.77	1.4e+03	501	! DNA encoding a Staphylococcus a	N_Geneseq_36:V75238	+	6.00	89.27	2.3e+03	809	! Staphylococcus aureus contig
N_Geneseq_36:Q12239	6.00	92.72	1.5e+03	504	! Clone 164/137 encoding PT-NABNR	N_Geneseq_36:V69440	+	6.00	89.27	2.3e+03	809	! Banana fruit ripening-relate
N_Geneseq_36:Q61601	6.00	92.55	1.5e+03	516	! AK beta subunit T30A gene fragm	N_Geneseq_36:V69471	-	6.00	89.26	2.3e+03	810	! Banana fruit ripening-relate
N_Geneseq_36:Q61602	6.00	92.55	1.5e+03	516	! AK wild type beta subunit gene	N_Geneseq_36:V47532	-	6.00	89.25	2.3e+03	811	! cDNA encoding a DAN and b57
N_Geneseq_36:V90747	6.00	92.52	1.5e+03	518	! Nucleotide sequence of clone z4	N_Geneseq_36:V26252	-	6.00	89.18	2.3e+03	819	! Photosynthetic organism car
N_Geneseq_36:N30847	6.00	92.48	1.5e+03	521	! Streptococcus pneumoniae	N_Geneseq_36:V20008	-	6.00	89.13	2.3e+03	825	! Enterococcus faecalis gene E
N_Geneseq_36:V88189	6.00	92.44	1.5e+03	524	! EST clone FZ428. New polynucleo	N_Geneseq_36:X30633	+	6.00	89.10	2.3e+03	828	! H. pylori cytoplasmic protei
N_Geneseq_36:V75897	6.00	92.32	1.5e+03	533	! Staphylococcus aureus contig S	N_Geneseq_36:X30634	+	6.00	89.10	2.3e+03	837	! H. pylori cytoplasmic protei
N_Geneseq_36:V27909	6.00	92.30	1.5e+03	534	! Human beta globin mutant gene	N_Geneseq_36:T11323	+	6.00	88.95	2.4e+03	845	! DNA encoding recombinant tra
N_Geneseq_36:V74668	6.00	92.30	1.5e+03	534	! Staphylococcus aureus contig S	N_Geneseq_36:Q26710	-	6.00	88.95	2.4e+03	846	! Gene regulating lipase expre
N_Geneseq_36:N29103	6.00	92.27	1.5e+03	536	! Human beta globin mutant gene	N_Geneseq_36:Q27267	-	6.00	88.90	2.4e+03	851	! Human FCER1 alpha-subunit an
N_Geneseq_36:N29104	6.00	92.27	1.5e+03	536	! Human beta globin mutant gene	N_Geneseq_36:X39982	+	6.00	88.89	2.4e+03	852	! Prostate cancer associated g
N_Geneseq_36:N29089	6.00	92.27	1.5e+03	536	! Human beta globin gene fragmen	N_Geneseq_36:X58030	+	6.00	88.84	2.4e+03	858	! Mucin MUC-1 gene promoter re
N_Geneseq_36:N29091	6.00	92.27	1.5e+03	536	! Human beta globin mutant gene	N_Geneseq_36:Q79601	+	6.00	88.84	2.4e+03	858	! Human MUC1 gene 5' sequence
N_Geneseq_36:N29092	6.00	92.27	1.5e+03	536	! Human beta globin mutant gene	N_Geneseq_36:N80471	+	6.00	88.74	2.4e+03	870	! Sequence specifically expre
N_Geneseq_36:V59519	6.00	92.09	1.6e+03	550	! Human secreted protein gene 9	N_Geneseq_36:V06105	-	6.00	88.63	2.5e+03	883	! Viral infection gene SEQ ID
N_Geneseq_36:N16967	6.00	92.07	1.6e+03	551	! Nucleic acid sequence of the sp	N_Geneseq_36:V45248	+	6.00	88.62	2.5e+03	885	! Mycobacterium tuberculosis 8
N_Geneseq_36:N14009	6.00	92.02	1.6e+03	555	! LEHR-6 coding sequence. Isolati	N_Geneseq_36:X39977	-	6.00	88.62	2.5e+03	885	! Prostate cancer associated g
N_Geneseq_36:N32700	6.00	92.02	1.6e+03	555	! LEHR-6 coding sequence. New iso	N_Geneseq_36:V06100	-	6.00	88.59	2.5e+03	888	! Viral infection gene SEQ ID
N_Geneseq_36:N32761	6.00	92.02	1.6e+03	555	! Murine ILK-6 polypeptide encod	N_Geneseq_36:V17184	+	6.00	88.56	2.5e+03	892	! cDNA encoding a novel human
N_Geneseq_36:X54633	6.00	91.93	1.6e+03	562	! Human ILK-4 receptor antisense c	N_Geneseq_36:V193661	+	6.00	88.49	2.5e+03	900	! Human secreted protein gene
N_Geneseq_36:X30894	6.00	91.89	1.6e+03	565	! Streptococcus pneumoniae genom	N_Geneseq_36:V06126	-	6.00	88.37	2.5e+03	915	! Viral infection gene SEQ ID
N_Geneseq_36:N78869	6.00	91.83	1.6e+03	570	! EST clone DI303. New polynucleo	N_Geneseq_36:N80315	+	6.00	88.33	2.6e+03	920	! Sequence

N_Geneseq_36:T85615	6.00	86.41	3.3e+03	1197	Alpha subunit of human high af	N_Geneseq_36:V24979	6.00	85.07	3.9e+03	1440	H. pylori secreted protein
N_Geneseq_36:V34204	6.00	86.41	3.3e+03	1197	Human secreted protein gene 51	N_Geneseq_36:V25102	6.00	85.07	3.9e+03	1440	H. pylori secreted protein
N_Geneseq_36:V36343	6.00	86.41	3.3e+03	1198	CDNA encoding the alpha chain	N_Geneseq_36:V25111	6.00	85.07	3.9e+03	1440	H. pylori secreted protein
N_Geneseq_36:Q14736	6.00	86.40	3.3e+03	1199	Human Fc(epsilon)RI alpha cDNA	N_Geneseq_36:Q61599	6.00	85.05	3.9e+03	1443	AK alpha subunit T279A gene
N_Geneseq_36:V45211	6.00	86.40	3.3e+03	1199	Human endosulfine B gene. Nove	N_Geneseq_36:Q61600	6.00	85.04	3.9e+03	1443	AK wild type alpha subunit
N_Geneseq_36:V50913	6.00	86.40	3.3e+03	1200	Nucleotide sequence of cluster	N_Geneseq_36:T93678	6.00	85.04	3.9e+03	1446	MLTL glycosidase 296 encod
N_Geneseq_36:X01143	6.00	86.40	3.3e+03	1200	M. tuberculosis antigen clone	N_Geneseq_36:T93697	6.00	85.04	3.9e+03	1446	Desulfurococcus MTL glyco
N_Geneseq_36:X01177	6.00	86.40	3.3e+03	1200	M. tuberculosis antigen clone	N_Geneseq_36:Q06188	6.00	85.03	3.9e+03	1448	Human Bone Morphogenesis P
N_Geneseq_36:Q27378	6.00	86.33	3.3e+03	1211	85-C. Recombinant peptide(s) a	N_Geneseq_36:Q32856	6.00	85.03	3.9e+03	1448	BMP7. Synergistic compen. f
N_Geneseq_36:N60836	6.00	86.31	3.3e+03	1215	5' region of bacterial expressi	N_Geneseq_36:Q13135	6.00	85.03	3.9e+03	1448	Human BMP-7 gene. Recombina
N_Geneseq_36:Q72693	6.00	86.31	3.3e+03	1215	Feedback inhibition release as	N_Geneseq_36:T98605	6.00	85.03	3.9e+03	1448	DNA encoding a S. pneumonia
N_Geneseq_36:Q72694	6.00	86.31	3.3e+03	1215	Feedback inhibition release as	N_Geneseq_36:T98605	6.00	85.03	3.9e+03	1448	DNA encoding a S. pneumonia
N_Geneseq_36:X00620	6.00	86.31	3.3e+03	1215	Human secreted protein gene 10	N_Geneseq_36:T32482	6.00	84.99	3.9e+03	1448	Transforming growth factor
N_Geneseq_36:T97223	6.00	86.29	3.3e+03	1217	Host integration factor himd 9	N_Geneseq_36:T31552	6.00	84.98	3.9e+03	1456	Human ice-3 homolog-1
N_Geneseq_36:V53300	6.00	86.25	3.3e+03	1224	DNA encoding a S. pneumoniae	N_Geneseq_36:T90595	6.00	84.93	3.9e+03	1457	Cell death protein ICH-11 e
N_Geneseq_36:V53308	6.00	86.23	3.3e+03	1227	D. immitis ankyrin nDIAK1227	N_Geneseq_36:Q14135	6.00	84.93	4.0e+03	1468	Enterococcus faecalis genom
N_Geneseq_36:V63009	6.00	86.23	3.3e+03	1227	D. immitis ankyrin nDIAK1227	N_Geneseq_36:Q14857	6.00	84.90	4.0e+03	1473	Fusarium oxysporum DSM 2672
N_Geneseq_36:V63006	6.00	86.23	3.3e+03	1228	D. immitis ankyrin nDIAK1228	N_Geneseq_36:Q26407	6.00	84.90	4.0e+03	1473	Fusarium oxysporum DSM 2672
N_Geneseq_36:V63007	6.00	86.23	3.3e+03	1228	D. immitis ankyrin nDIAK1228	N_Geneseq_36:Q26407	6.00	84.90	4.0e+03	1473	Fusarium oxysporum DSM 2672
N_Geneseq_36:V63310	6.00	86.23	3.3e+03	1228	D. immitis ankyrin nDIAK1228	N_Geneseq_36:Q26407	6.00	84.90	4.0e+03	1473	Fusarium oxysporum DSM 2672
N_Geneseq_36:X04357	6.00	86.16	3.4e+03	1239	Human secreted protein gene 47	N_Geneseq_36:Q25933	6.00	84.90	4.0e+03	1473	Cellulase contained in a de
N_Geneseq_36:X39976	6.00	86.16	3.4e+03	1240	Prostate cancer associated gen	N_Geneseq_36:Q41733	6.00	84.90	4.0e+03	1473	Endoglucanase gene. Laundry
N_Geneseq_36:V13114	6.00	86.13	3.4e+03	1245	Satellite tobacco necrosis vir	N_Geneseq_36:Q41942	6.00	84.90	4.0e+03	1473	Dye transfer inhibiting com
N_Geneseq_36:V52117	6.00	86.09	3.4e+03	1251	H. pylori cell envelope OMP OR	N_Geneseq_36:V16103	6.00	84.90	4.0e+03	1473	Endoglucanase enzyme. Deter
N_Geneseq_36:V59575	6.00	86.07	3.4e+03	1251	Human secreted protein gene 65	N_Geneseq_36:T71068	6.00	84.85	4.0e+03	1473	Fusarium oxysporum surface
N_Geneseq_36:T18905	6.00	86.05	3.4e+03	1258	Delta 6-palmitoyl acyl carrier	N_Geneseq_36:T60296	6.00	84.84	4.0e+03	1482	Sequence encoding equine in
N_Geneseq_36:Q48165	6.00	86.02	3.4e+03	1263	B. flavum AK gene. Release DNA	N_Geneseq_36:X00698	6.00	84.83	4.0e+03	1486	Human interleukin-6 nucleot
N_Geneseq_36:Q72692	6.00	86.02	3.4e+03	1263	Feedback inhibition release as	N_Geneseq_36:T98639	6.00	84.83	4.0e+03	1487	Human secreted protein gene
N_Geneseq_36:Q72695	6.00	86.02	3.4e+03	1263	Feedback inhibition release as	N_Geneseq_36:T98639	6.00	84.83	4.0e+03	1488	Thermolabile glucose-6-phos
N_Geneseq_36:Q72696	6.00	86.02	3.4e+03	1263	Feedback inhibition release as	N_Geneseq_36:T98639	6.00	84.82	4.0e+03	1490	Methods for diagnosing Frie
N_Geneseq_36:Q72697	6.00	86.02	3.4e+03	1263	Feedback inhibition release as	N_Geneseq_36:T98639	6.00	84.82	4.0e+03	1490	Methods for diagnosing Frie
N_Geneseq_36:T71310	6.00	86.02	3.4e+03	1263	DNA encoding histidyl-tRNA syn	N_Geneseq_36:T36148	6.00	84.77	4.0e+03	1499	Sequence of ptz gene found
N_Geneseq_36:T13486	6.00	85.97	3.5e+03	1272	Rice NADH-dependent reductase	N_Geneseq_36:Q21726	6.00	84.77	4.0e+03	1500	Human homolog of murine I
N_Geneseq_36:Q53148	6.00	85.94	3.5e+03	1277	Sequence encoding osteogenic f	N_Geneseq_36:T89405	6.00	84.76	4.0e+03	1501	Parainfluenza type 4A virus
N_Geneseq_36:Q72714	6.00	85.94	3.5e+03	1277	Osteogenic protein 1D OP1D fus	N_Geneseq_36:T89405	6.00	84.76	4.0e+03	1501	Methods for diagnosing Frie
N_Geneseq_36:T02608	6.00	85.94	3.5e+03	1277	OP1(d) cDNA encoding OP1D fus	N_Geneseq_36:Q53147	6.00	84.76	4.0e+03	1501	Methods for diagnosing Frie
N_Geneseq_36:V15211	6.00	85.94	3.5e+03	1277	Human osteogenic fusion protel	N_Geneseq_36:Q72713	6.00	84.74	4.0e+03	1505	Osteogenic protein 1C OP1C
N_Geneseq_36:X00236	6.00	85.94	3.5e+03	1277	Osteogenic fusion protein OP1D	N_Geneseq_36:T02607	6.00	84.74	4.0e+03	1505	OP1(c) cDNA encoding OP1C f
N_Geneseq_36:T77317	6.00	85.89	3.5e+03	1287	Class II EPSPS DNA for glyphos	N_Geneseq_36:V15210	6.00	84.74	4.0e+03	1505	Human osteogenic fusion pro
N_Geneseq_36:T77393	6.00	85.89	3.5e+03	1287	Class II EPSPS DNA for glyphos	N_Geneseq_36:X00235	6.00	84.74	4.0e+03	1505	Osteogenic fusion protein o
N_Geneseq_36:V58018	6.00	85.89	3.5e+03	1287	Bacillus subtilis Class II EPS	N_Geneseq_36:Q59143	6.00	84.74	4.0e+03	1506	DM beta cDNA. Identifying m
N_Geneseq_36:X37458	6.00	85.89	3.5e+03	1287	Human secreted protein cDNA fr	N_Geneseq_36:T93628	6.00	84.74	4.0e+03	1506	HPV6/11 hybrid DNA encoding
N_Geneseq_36:T73325	6.00	85.87	3.5e+03	1290	Human CD18 cDNA. Oligonucleoti	N_Geneseq_36:V16721	6.00	84.74	4.0e+03	1506	Nucleic acid encoding rat N
N_Geneseq_36:T73326	6.00	85.87	3.5e+03	1290	Human CD18 pseudogene clone 6	N_Geneseq_36:V33585	6.00	84.74	4.0e+03	1506	Human papillomavirus 6/11 h
N_Geneseq_36:Q21967	6.00	85.84	3.5e+03	1295	HindIII-NruI restriction fragm	N_Geneseq_36:N81125	6.00	84.72	4.1e+03	1509	Deacetoxycyclosporin-C-sy
N_Geneseq_36:Q49735	6.00	85.83	3.5e+03	1296	OP-1 gene. Producing a DNA seq	N_Geneseq_36:Q53470	6.00	84.72	4.1e+03	1509	PTM gene B1. Expression of
N_Geneseq_36:T93789	6.00	85.80	3.5e+03	1302	Mouse SH3P7 gene. Identifying	N_Geneseq_36:V59788	6.00	84.69	4.1e+03	1517	Human secreted protein gene
N_Geneseq_36:V61795	6.00	85.80	3.5e+03	1302	Human betac cytoplasmic domain	N_Geneseq_36:V16720	6.00	84.68	4.1e+03	1518	Nucleic acid encoding human
N_Geneseq_36:Q44450	6.00	85.79	3.5e+03	1303	Cysteine synthase. Novel gene	N_Geneseq_36:V20072	6.00	84.68	4.1e+03	1518	N. Wiskott-Aldrich syndrome
N_Geneseq_36:V34225	6.00	85.68	3.6e+03	1324	Human secreted protein gene 4	N_Geneseq_36:T34595	6.00	84.68	4.1e+03	1518	M. vaccae pota homologue en
N_Geneseq_36:X22281	6.00	85.66	3.6e+03	1327	Human secreted protein gene 63	N_Geneseq_36:Q02852	6.00	84.68	4.1e+03	1519	DNA of genomic clone encodi
N_Geneseq_36:V30803	6.00	85.66	3.6e+03	1328	Human wild-type prohibitin gen	N_Geneseq_36:T22273	6.00	84.66	4.1e+03	1523	Human secreted protein gene
N_Geneseq_36:X30745	6.00	85.63	3.6e+03	1332	Streptococcus pneumoniae genom	N_Geneseq_36:T40119	6.00	84.65	4.1e+03	1523	Human papillomavirus type 1
N_Geneseq_36:X17775	6.00	85.60	3.6e+03	1338	HTLAK33 encoding cDNA (EST der	N_Geneseq_36:T85928	6.00	84.64	4.1e+03	1527	DNA encoding lysophosphatid
N_Geneseq_36:X30759	6.00	85.59	3.6e+03	1340	Streptococcus pneumoniae genom	N_Geneseq_36:T05577	6.00	84.63	4.1e+03	1528	Human p69 cDNA. Antigen com
N_Geneseq_36:V59747	6.00	85.55	3.6e+03	1348	Human secreted protein gene 89	N_Geneseq_36:V44370	6.00	84.58	4.1e+03	1539	Mycobacterium tuberculosis
N_Geneseq_36:X14148	6.00	85.54	3.7e+03	1350	H. pylori GHPO 422 gene. New l	N_Geneseq_36:V64478	6.00	84.58	4.1e+03	1539	M. tuberculosis immunogeni
N_Geneseq_36:X13925	6.00	85.50	3.7e+03	1357	H. pylori GHPO 541 gene. New l	N_Geneseq_36:V72588	6.00	84.57	4.1e+03	1542	Streptococcus pneumoniae pr
N_Geneseq_36:V52382	6.00	85.46	3.7e+03	1364	Streptococcus pneumoniae genom	N_Geneseq_36:X14075	6.00	84.53	4.2e+03	1550	H. pylori GHPO 1102 gene. N
N_Geneseq_36:X04848	6.00	85.41	3.7e+03	1373	Human lysophosphatidic acid ac	N_Geneseq_36:T20313	6.00	84.52	4.2e+03	1552	Borrelia burgdorferi polyu
N_Geneseq_36:X30167	6.00	85.40	3.7e+03	1376	Human secreted protein gene 23	N_Geneseq_36:T47652	6.00	84.45	4.2e+03	1565	DNA encoding transforming
N_Geneseq_36:V34157	6.00	85.39	3.7e+03	1378	Human secreted protein gene 4	N_Geneseq_36:Q12773	6.00	84.45	4.2e+03	1567	Human polymorphic epithelia
N_Geneseq_36:X37485	6.00	85.32	3.8e+03	1391	Human secreted protein cDNA fr	N_Geneseq_36:T29199	6.00	84.33	4.3e+03	1593	Rat acetylglucosaminyl tran
N_Geneseq_36:T72242	6.00	85.31	3.8e+03	1398	DNA encoding modified OP-1. Tr	N_Geneseq_36:T77094	6.00	84.30	4.3e+03	1599	Human N-acetylglucosaminyl
N_Geneseq_36:Q27816	6.00	85.28	3.8e+03	1398	Adrenalin receptor subtype bet	N_Geneseq_36:T43204	6.00	84.29	4.3e+03	1602	Human mitogen-activated pro
N_Geneseq_36:T66442	6.00	85.27	3.8e+03	1401	Midase gene. Nucleic acid enc	N_Geneseq_36:X07066	6.00	84.29	4.3e+03	1602	Human mitogen-activated pro
N_Geneseq_36:X32491	6.00	85.23	3.8e+03	1407	Seq ID NC.6 used in a method	N_Geneseq_36:Q26528	6.00	84.28	4.3e+03	1603	Acyl-ACP thioesterase. Modl
N_Geneseq_36:Q80524	6.00	85.22	3.8e+03	1410	Oxidoreducing avermectin DNA	N_Geneseq_36:Q26814	6.00	84.25	4.3e+03	1611	IGFBP6 DNA. Insulin-like gr
N_Geneseq_36:T15579	6.00	85.19	3.8e+03	1416	Human secreted protein gene 89	N_Geneseq_36:T26196	6.00	84.23	4.3e+03	1611	DNA sequence of glycosyltra
N_Geneseq_36:Q44494	6.00	85.12	3.9e+03	1430	ICE-like apoptosis protease-2	N_Geneseq_36:Q67875	6.00	84.23	4.3e+03	1615	C6 insertion plasmid pC6L
N_Geneseq_36:Q44494	6.00	85.11	3.9e+03	1432	Bacillus coagulans proline imi	N_Geneseq_36:T04704	6.00	84.23	4.3e+03	1615	ALVAC C6 flanking arm used
N_Geneseq_36:V24978	6.00	85.07	3.9e+03	1440	H. pylori secreted protein. ORF	N_Geneseq_36:V26972	6.00	84.23	4.3e+03	1615	Canarypox virus ALVAC pC6L

N_Geneseq_36.V27175	6.00	84.18	4.3e+03	1626	G-protein coupled receptor, 1c	N_Geneseq_36.Q38944	6.00	83.35	4.8e+03	1822	Human osteogenic protein 1
N_Geneseq_36.Q54496	6.00	84.17	4.4e+03	1629	Alcaligenes bronchisepticus ad	N_Geneseq_36.Q38733	6.00	83.35	4.8e+03	1822	Human OP-1. Morphogen-induc
N_Geneseq_36.N40071	6.00	84.14	4.4e+03	1635	Sequence coding for Herpes sim	N_Geneseq_36.Q38857	6.00	83.35	4.8e+03	1822	Morphogen hOP1 coding sequ
N_Geneseq_36.T51325	6.00	84.14	4.4e+03	1635	HSV glycoprotein D-2 DNA. Vari	N_Geneseq_36.Q53141	6.00	83.35	4.8e+03	1822	Sequence encoding human ost
N_Geneseq_36.V31670	6.00	84.14	4.4e+03	1635	Herpes simplex virus type 2 gl	N_Geneseq_36.Q51918	6.00	83.35	4.8e+03	1822	hOP1 cDNA. Morphogenic prot
N_Geneseq_36.Q10429	6.00	84.13	4.4e+03	1637	Male hormone-dependent gene. N	N_Geneseq_36.Q57916	6.00	83.35	4.8e+03	1822	Human osteogenic protein ho
N_Geneseq_36.V80562	6.00	84.12	4.4e+03	1639	Human lysophosphatidic acid ac	N_Geneseq_36.Q67311	6.00	83.35	4.8e+03	1822	Human OP-1. Morphogen-induc
N_Geneseq_36.Q14835	6.00	84.11	4.4e+03	1642	Human serotonin receptor-encod	N_Geneseq_36.Q04511	6.00	83.35	4.8e+03	1822	Human OP-1. Maintaining int
N_Geneseq_36.Q61596	6.00	84.10	4.4e+03	1643	AK wild type alpha subunit gen	N_Geneseq_36.Q65391	6.00	83.35	4.8e+03	1822	Osteogenic protein hOP1-PP.
N_Geneseq_36.Q61597	6.00	84.10	4.4e+03	1643	AK beta subunit T30A gene. The	N_Geneseq_36.Q54162	6.00	83.35	4.8e+03	1822	Human OP-1. Use of morphoge
N_Geneseq_36.Q61595	6.00	84.10	4.4e+03	1643	AK wild type beta subunit gene	N_Geneseq_36.Q54755	6.00	83.35	4.8e+03	1822	Osteogenic protein OPl. Bon
N_Geneseq_36.Q84970	6.00	84.10	4.4e+03	1643	AK alpha subunit r279A gene. P	N_Geneseq_36.Q71425	6.00	83.35	4.8e+03	1822	Osteogenic protein OPl. cDNA
N_Geneseq_36.Q84971	6.00	84.10	4.4e+03	1643	C. glutamicum gene. Variant of	N_Geneseq_36.Q72703	6.00	83.35	4.8e+03	1822	hOP1 human osteogenic prote
N_Geneseq_36.T03326	6.00	84.10	4.4e+03	1643	Corynebacterium glutamicum asp	N_Geneseq_36.T06038	6.00	83.35	4.8e+03	1822	Human osteogenic protein 1
N_Geneseq_36.T49288	6.00	84.10	4.4e+03	1643	Fragment of the B. lactofermen	N_Geneseq_36.T02597	6.00	83.35	4.8e+03	1822	hOP-1 cDNA. Antibodies with
N_Geneseq_36.T49289	6.00	84.10	4.4e+03	1643	DNA encoding B. lactofermentu	N_Geneseq_36.T06771	6.00	83.35	4.8e+03	1822	Human osteogenic protein Op
N_Geneseq_36.T49289	6.00	84.10	4.4e+03	1643	DNA encoding B. lactofermentu	N_Geneseq_36.T33441	6.00	83.35	4.8e+03	1822	Human osteogenic protein Op
N_Geneseq_36.T49289	6.00	84.10	4.4e+03	1643	Brevibacterium lactofermentu	N_Geneseq_36.T73207	6.00	83.35	4.8e+03	1822	Human hippocampal osteogeni
N_Geneseq_36.V15781	6.00	84.10	4.4e+03	1643	B. lactofermentum aspartokinase	N_Geneseq_36.T87879	6.00	83.35	4.8e+03	1822	Human osteogenic protein 1
N_Geneseq_36.V15782	6.00	84.10	4.4e+03	1643	B. lactofermentum aspartokinase	N_Geneseq_36.T93810	6.00	83.35	4.8e+03	1822	cDNA encoding human osteoge
N_Geneseq_36.V40254	6.00	84.10	4.4e+03	1643	Brevibacterium lactofermentu	N_Geneseq_36.T93810	6.00	83.35	4.8e+03	1822	Human osteogenic protein-1
N_Geneseq_36.V40255	6.00	84.10	4.4e+03	1643	Brevibacterium lactofermentu	N_Geneseq_36.V10345	6.00	83.35	4.8e+03	1822	Human OP-1 cDNA. Treatment
N_Geneseq_36.V41300	6.00	84.10	4.4e+03	1643	Aspartokinase lysc gene. New	N_Geneseq_36.V15204	6.00	83.35	4.8e+03	1822	Human osteogenic protein Op
N_Geneseq_36.V41316	6.00	84.10	4.4e+03	1643	Mutant aspartokinase lysc gene	N_Geneseq_36.V19533	6.00	83.35	4.8e+03	1822	Human osteogenic protein (h
N_Geneseq_36.Q65519	6.00	84.07	4.4e+03	1650	Human insulin-like growth fact	N_Geneseq_36.V33583	6.00	83.35	4.8e+03	1822	Human osteogenic protein 1
N_Geneseq_36.T72216	6.00	84.07	4.4e+03	1650	Grapevine leafroll virus hsp-7	N_Geneseq_36.V41572	6.00	83.35	4.8e+03	1822	Morphogenically active oest
N_Geneseq_36.Q80218	6.00	84.07	4.4e+03	1651	Human NDF-alpha2b clone 17 DNA	N_Geneseq_36.V60227	6.00	83.35	4.8e+03	1822	Nucleotide sequence encodin
N_Geneseq_36.Q66990	6.00	84.05	4.4e+03	1656	DF3 enhancer. Isolated DF3 en	N_Geneseq_36.V65199	6.00	83.35	4.8e+03	1822	Human osteogenic protein CD
N_Geneseq_36.V77543	6.00	84.05	4.4e+03	1656	Staphylococcus aureus contig	N_Geneseq_36.V80656	6.00	83.35	4.8e+03	1822	Human osteogenic protein Op
N_Geneseq_36.T45050	6.00	84.00	4.4e+03	1666	Beta-globin gene for retrovira	N_Geneseq_36.V80735	6.00	83.35	4.8e+03	1822	Human osteogenic protein Op
N_Geneseq_36.T70926	6.00	83.94	4.5e+03	1680	Plasmid pTUB616 encoding heat	N_Geneseq_36.X00229	6.00	83.35	4.8e+03	1822	Human osteogenic protein Op
N_Geneseq_36.V34545	6.00	83.94	4.5e+03	1680	Saccharomyces cerevisiae RNA	N_Geneseq_36.X02631	6.00	83.35	4.8e+03	1823	hOP1 cDNA. Use morphogens t
N_Geneseq_36.Q54134	6.00	83.92	4.5e+03	1685	Sequence encoding mouse adipoc	N_Geneseq_36.Q87837	6.00	83.35	4.8e+03	1823	Human calcium channel subun
N_Geneseq_36.T44455	6.00	83.92	4.5e+03	1685	cDNA encoding murine p154. Mam	N_Geneseq_36.V16293	6.00	83.33	4.8e+03	1826	Tomato phosphate transporte
N_Geneseq_36.T08083	6.00	83.90	4.5e+03	1699	Alpha-1,3/4-fucosidase gene. A	N_Geneseq_36.Q14627	6.00	83.32	4.9e+03	1839	Human Glioblastoma cell cDN
N_Geneseq_36.T05038	6.00	83.85	4.5e+03	1700	Sequence encoding alpha-amyli	N_Geneseq_36.T33467	6.00	83.32	4.9e+03	1829	Plasmid pJC265 (ATCC 68598)
N_Geneseq_36.T04588	6.00	83.79	4.6e+03	1715	Rat Vhh-1 cDNA. Nucleic acid	N_Geneseq_36.T33630	6.00	83.29	4.9e+03	1836	Rat p65 oncofetal protein
N_Geneseq_36.Q92299	6.00	83.78	4.6e+03	1717	Mango class II thioesterase MA	N_Geneseq_36.V53631	6.00	83.27	4.9e+03	1841	Pyranose oxidase encoding s
N_Geneseq_36.V15233	6.00	83.74	4.6e+03	1717	Class II acyl-ACP thioesterase	N_Geneseq_36.T34420	6.00	83.16	5.0e+03	1869	Thermococcus amidase gene.
N_Geneseq_36.T70920	6.00	83.77	4.6e+03	1719	Plasmid pTUB613 encoding heat	N_Geneseq_36.V12887	6.00	83.16	5.0e+03	1869	Thermococcus amidase gene.
N_Geneseq_36.Q66988	6.00	83.75	4.6e+03	1725	5' flanking region oof the DF3	N_Geneseq_36.Q90623	6.00	83.15	5.0e+03	1872	Murine osteogenic protein (p
N_Geneseq_36.N90523	6.00	83.74	4.6e+03	1726	Herpes Simplex virus-2 gd gene	N_Geneseq_36.Q24517	6.00	83.15	5.0e+03	1872	Murine osteogenic protein Op
N_Geneseq_36.N50488	6.00	83.73	4.6e+03	1730	Sequence of herpes simplex vir	N_Geneseq_36.Q28736	6.00	83.15	5.0e+03	1873	Murine osteogenic protein m
N_Geneseq_36.X23754	6.00	83.71	4.6e+03	1734	Rice anthranilate synthase fir	N_Geneseq_36.Q38945	6.00	83.15	5.0e+03	1873	Murine osteogenic protein 1
N_Geneseq_36.X23748	6.00	83.67	4.6e+03	1734	Rice anthranilate synthase fir	N_Geneseq_36.Q38734	6.00	83.15	5.0e+03	1873	Murine pro-Op-1. Morphogen-
N_Geneseq_36.T85939	6.00	83.67	4.6e+03	1743	Shigella flexneri IpaB gene. S	N_Geneseq_36.Q38858	6.00	83.15	5.0e+03	1873	Morphogen mOP1. Coding sequ
N_Geneseq_36.T92839	6.00	83.61	4.7e+03	1758	Flea saliva protein fspM(F) cD	N_Geneseq_36.Q51153	6.00	83.15	5.0e+03	1873	Sequence encoding murine os
N_Geneseq_36.V73387	6.00	83.61	4.7e+03	1758	Flea saliva protein nucleic ac	N_Geneseq_36.Q51199	6.00	83.15	5.0e+03	1873	mOP1 cDNA. Morphogenic prot
N_Geneseq_36.Q13771	6.00	83.59	4.7e+03	1763	Human polymorphic epithelial m	N_Geneseq_36.Q58232	6.00	83.15	5.0e+03	1873	mOP1 cDNA. Use morphogens t
N_Geneseq_36.Q79143	6.00	83.58	4.7e+03	1765	Hepatitis C virus gene SR037-5	N_Geneseq_36.Q58051	6.00	83.15	5.0e+03	1873	Mouse osteogenic protein mO
N_Geneseq_36.V59594	6.00	83.54	4.7e+03	1774	Human secreted protein gene 84	N_Geneseq_36.Q67312	6.00	83.15	5.0e+03	1873	Murine OP-1. Morphogen-indu
N_Geneseq_36.V47531	6.00	83.52	4.7e+03	1781	cDNA encoding a DAN and b57 pr	N_Geneseq_36.Q65117	6.00	83.15	5.0e+03	1873	Murine OP-1. Maintaining in
N_Geneseq_36.Q42191	6.00	83.50	4.7e+03	1785	PM-1 antigen. Neuro-endocrine	N_Geneseq_36.Q65392	6.00	83.15	5.0e+03	1873	Osteogenic protein mOP1-pp.
N_Geneseq_36.T04154	6.00	83.47	4.8e+03	1791	BCG DapB gene. Gene involved	N_Geneseq_36.Q45163	6.00	83.15	5.0e+03	1873	Murine OP-1. Use of morphog
N_Geneseq_36.T03250	6.00	83.46	4.8e+03	1794	Enterobacter sp. S262 sucrose	N_Geneseq_36.Q27204	6.00	83.15	5.0e+03	1873	mOP1-PP prepro form mouse o
N_Geneseq_36.V22703	6.00	83.45	4.8e+03	1797	Human recombinase hSRC2 cDNA.	N_Geneseq_36.Q80220	6.00	83.15	5.0e+03	1873	Human NDF-beta1a clone P13
N_Geneseq_36.Q05999	6.00	83.44	4.8e+03	1800	Sequence encoding foetal human	N_Geneseq_36.T92598	6.00	83.15	5.0e+03	1873	mOP-1 cDNA. Antibodies with
N_Geneseq_36.T90412	6.00	83.44	4.8e+03	1800	Human glutamate transporter pr	N_Geneseq_36.T97879	6.00	83.15	5.0e+03	1873	cDNA encoding mouse osteoge
N_Geneseq_36.V61147	6.00	83.44	4.8e+03	1800	Human excitatory amino acid tr	N_Geneseq_36.V10346	6.00	83.15	5.0e+03	1873	Mouse OP-1 cDNA. Treatment
N_Geneseq_36.V61177	6.00	83.44	4.8e+03	1800	Human secreted protein bdl64-7	N_Geneseq_36.V15216	6.00	83.15	5.0e+03	1873	Mouse osteogenic protein Op
N_Geneseq_36.V62331	6.00	83.44	4.8e+03	1800	Human excitatory amino acid tr	N_Geneseq_36.V19534	6.00	83.15	5.0e+03	1873	Mouse osteogenic protein (m
N_Geneseq_36.T03249	6.00	83.43	4.8e+03	1803	Protaminobacter rubrum sucrose	N_Geneseq_36.V32584	6.00	83.15	5.0e+03	1873	Murine osteogenic protein 1
N_Geneseq_36.Q04673	6.00	83.42	4.8e+03	1804	Sul gene containing fragment (1	N_Geneseq_36.X00241	6.00	83.15	5.0e+03	1873	Murine osteogenic protein M
N_Geneseq_36.T15008	6.00	83.40	4.8e+03	1809	Mouse Elf-1 cDNA. Murine and c	N_Geneseq_36.V33033	6.00	83.14	5.0e+03	1875	Dirofilaria immitis transgl
N_Geneseq_36.Q88068	6.00	83.38	4.8e+03	1814	Bacillus stearothermophilus ad	N_Geneseq_36.T61117	6.00	83.13	5.0e+03	1878	DNA encoding OP-1. Transfor
N_Geneseq_36.Q93033	6.00	83.38	4.8e+03	1814	Bacillus stearothermophilus al	N_Geneseq_36.T03246	6.00	83.08	5.0e+03	1890	Streptococcus rubrum suc
N_Geneseq_36.V02473	6.00	83.38	4.8e+03	1814	Bacillus stearothermophilus Te	N_Geneseq_36.V52352	6.00	83.04	5.0e+03	1900	Streptococcus pneumoniae ge
N_Geneseq_36.Q04731	6.00	83.36	4.8e+03	1820	cDNA sequence from mRNA of gl	N_Geneseq_36.V39012	6.00	83.02	5.0e+03	1906	CD30 ligand gene used in th
N_Geneseq_36.Q11803	6.00	83.35	4.8e+03	1822	Sequence encoding mammalian os	N_Geneseq_36.V11055	6.00	82.97	5.1e+03	1919	Clone 19 encoding murine IL
N_Geneseq_36.Q24518	6.00	83.35	4.8e+03	1822	Human osteogenic protein hOP1	N_Geneseq_36.T31931	6.00	82.96	5.1e+03	1922	Retinoid X receptor interac
N_Geneseq_36.Q28735	6.00	83.35	4.8e+03	1822	Human osteogenic protein hOP1	N_Geneseq_36.V59546	6.00	82.95	5.1e+03	1924	Human secreted protein gene

N_Geneseq_36:X08426	6.00	82.93	5.1e+03	1929	Acidic leucine aminopeptidase	N_Geneseq_36:T47338	6.00	81.54	6.0e+03	2303	Murine developmentally-regu
N_Geneseq_36:Q04472	6.00	82.85	5.2e+03	1950	Human papilloma virus 52 clone	N_Geneseq_36:T96034	6.00	81.64	6.0e+03	2303	Rat kidney injury related m
N_Geneseq_36:Q56791	6.00	82.78	5.2e+03	1971	CDNA encoding receptor for C-t	N_Geneseq_36:X18506	6.00	81.64	6.0e+03	2303	Murine del-1 encoding cDNA.
N_Geneseq_36:T67025	6.00	82.78	5.2e+03	1971	Rat beta-amyloid precursor pro	N_Geneseq_36:Q39149	6.00	81.63	6.0e+03	2305	Annexin XI gene in clone pc
N_Geneseq_36:X08428	6.00	82.78	5.2e+03	1971	Acidic leucine aminopeptidase	N_Geneseq_36:T47339	6.00	81.62	6.0e+03	2308	Murine del-1 truncated mino
N_Geneseq_36:Q28118	6.00	82.73	5.2e+03	1983	Human norepinephrine transpor	N_Geneseq_36:X18507	6.00	81.62	6.0e+03	2308	Truncated murine del-1 enco
N_Geneseq_36:X33115	6.00	82.73	5.2e+03	1983	Human PRL-1 phosphatase encodi	N_Geneseq_36:Q39150	6.00	81.61	6.0e+03	2311	Annexin XI gene in clone pc
N_Geneseq_36:T69212	6.00	82.72	5.2e+03	1986	Human PRL-1 phosphatase encodi	N_Geneseq_36:Q82804	6.00	81.61	6.0e+03	2311	Murine inward rectifier K+ c
N_Geneseq_36:T90830	6.00	82.68	5.3e+03	1999	CDNA encoding polyurethane es	N_Geneseq_36:N90816	6.00	81.59	6.1e+03	2319	Membrane-bound aldehyde deh
N_Geneseq_36:Q05549	6.00	82.67	5.3e+03	1999	Sequence encoding Newcastle di	N_Geneseq_36:T32962	6.00	81.58	6.1e+03	2321	Mouse developmental kinase
N_Geneseq_36:N91000	6.00	82.67	5.3e+03	2000	Sequence encoding Newcastle di	N_Geneseq_36:X14459	6.00	81.56	6.1e+03	2322	Mouse developmental kinase
N_Geneseq_36:T68842	6.00	82.67	5.3e+03	2000	Sequence encoding Newcastle di	N_Geneseq_36:T688216	6.00	81.54	6.1e+03	2327	H. pylori GHP 961 gene. Ne
N_Geneseq_36:T68823	6.00	82.63	5.3e+03	2001	Mycobacterium tuberculosis sig	N_Geneseq_36:Q80216	6.00	81.47	6.2e+03	2335	Human prionDF-ALPHA2B DNA. N
N_Geneseq_36:N91804	6.00	82.61	5.3e+03	2012	Enterococcus faecalis genome c	N_Geneseq_36:Q80229	6.00	81.46	6.2e+03	2356	Rat NDF clone 38 DNA. New r
N_Geneseq_36:N19484	6.00	82.60	5.3e+03	2017	Cytosolic phospholipase A2/B c	N_Geneseq_36:T84986	6.00	81.47	6.2e+03	2361	BuMkholderia cepacia Insert
N_Geneseq_36:N90641	6.00	82.60	5.3e+03	2019	Human papilloma virus type 33	N_Geneseq_36:X04316	6.00	81.46	6.2e+03	2361	Human secreted protein gene
N_Geneseq_36:T16292	6.00	82.59	5.3e+03	2020	Staphylococcus aureus mecB ORF	N_Geneseq_36:Q55137	6.00	81.46	6.2e+03	2362	Staphylococcus epidermidis
N_Geneseq_36:Q31542	6.00	82.56	5.3e+03	2023	Tomato phosphatate transporter	N_Geneseq_36:Q77853	6.00	81.43	6.2e+03	2369	Mycoplasma gallisepticum an
N_Geneseq_36:X01163	6.00	82.52	5.4e+03	2029	Human heregulin beta-1 in clon	N_Geneseq_36:T63470	6.00	81.43	6.2e+03	2371	CDNA encoding a human chron
N_Geneseq_36:X01129	6.00	82.52	5.4e+03	2041	M. tuberculosis antigen clone	N_Geneseq_36:T26997	6.00	81.42	6.2e+03	2373	N. meningitidis serotype C
N_Geneseq_36:Q24216	6.00	82.47	5.4e+03	2054	M. tuberculosis antigen clone	N_Geneseq_36:T26998	6.00	81.41	6.2e+03	2376	N. meningitidis serotype A
N_Geneseq_36:T59236	6.00	82.47	5.4e+03	2054	Squalene synthase. Nucleic aci	N_Geneseq_36:T72590	6.00	81.41	6.2e+03	2376	Mouse neurotysin encoding
N_Geneseq_36:Q40093	6.00	82.45	5.4e+03	2060	S. cerevisiae squalene synthet	N_Geneseq_36:T67024	6.00	81.40	6.2e+03	2379	N. meningitidis serotype B
N_Geneseq_36:N90340	6.00	82.45	5.4e+03	2061	CPH gene. Isolated fungal dim	N_Geneseq_36:X17774	6.00	81.36	6.2e+03	2394	Rat beta-amyloid precursor
N_Geneseq_36:Q41746	6.00	82.43	5.4e+03	2066	Sequence encoding a receptor F	N_Geneseq_36:Q06622	6.00	81.34	6.2e+03	2400	YAK-1 related serine/threon
N_Geneseq_36:Q03288	6.00	82.40	5.5e+03	2076	IL-6 receptor coding sequence.	N_Geneseq_36:Q06622	6.00	81.25	6.3e+03	2400	Asparagine synthetase AS1 g
N_Geneseq_36:Q90719	6.00	82.38	5.5e+03	2081	Gene encoding yeast NADPH-cytd	N_Geneseq_36:Q04305	6.00	81.25	6.3e+03	2430	Coding region of murine IL-
N_Geneseq_36:Q90721	6.00	82.38	5.5e+03	2081	B. burgdorferi strain K48 anti	N_Geneseq_36:T61520	6.00	81.25	6.3e+03	2430	Rat NDF clone 42A DNA. New
N_Geneseq_36:Q88694	6.00	82.37	5.5e+03	2084	B. burgdorferi strain P10B an	N_Geneseq_36:T17555	6.00	81.25	6.3e+03	2430	Murine interleukin-4 recept
N_Geneseq_36:N81504	6.00	82.37	5.5e+03	2085	HindIII fragment pNF2009 (HA9)	N_Geneseq_36:X38332	6.00	81.25	6.3e+03	2430	Mouse interleukin-4 recept
N_Geneseq_36:N80847	6.00	82.36	5.5e+03	2087	OPPE 45 kDa subunit gene. New	N_Geneseq_36:X08858	6.00	81.25	6.3e+03	2430	Mouse interleukin-4 recept
N_Geneseq_36:Q67458	6.00	82.36	5.5e+03	2088	CDNA contg. region encoding BSF	N_Geneseq_36:T61456	6.00	81.17	6.4e+03	2457	C-proteinase clone pCP-1. N
N_Geneseq_36:T68825	6.00	82.28	5.6e+03	2109	Genes tau-protein kinase I (TF	N_Geneseq_36:Q13311	6.00	81.13	6.4e+03	2458	Methods for diagnosing Frie
N_Geneseq_36:Q05990	6.00	82.27	5.6e+03	2112	Cytosolic phospholipase A2/B c	N_Geneseq_36:Q13311	6.00	81.08	6.4e+03	2485	Human interleukin-1 beta co
N_Geneseq_36:T68826	6.00	82.26	5.6e+03	2115	Sequence encoding melanoma as	N_Geneseq_36:Q13311	6.00	81.07	6.4e+03	2487	BMPI. Synergistic compsn. f
N_Geneseq_36:X13605	6.00	82.25	5.6e+03	2116	Encodes PT-NANBH viral structu	N_Geneseq_36:Q13311	6.00	81.07	6.4e+03	2491	Human heregulin-beta2 clone
N_Geneseq_36:X05992	6.00	82.22	5.6e+03	2119	Enterococcus faecalis genome c	N_Geneseq_36:Q13311	6.00	81.07	6.4e+03	2491	Human interleukin-1 beta co
N_Geneseq_36:T96726	6.00	82.21	5.6e+03	2131	Kidney injury associated molec	N_Geneseq_36:Q13311	6.00	81.07	6.4e+03	2491	Human trident transcription
N_Geneseq_36:X01587	6.00	82.21	5.6e+03	2131	Melanoma-specific immunogen, B	N_Geneseq_36:N80631	6.00	81.04	6.5e+03	2500	Human Bone Morphogenic Prot
N_Geneseq_36:T15220	6.00	82.19	5.6e+03	2136	pMEL17 cDNA. Melanoma-specifi	N_Geneseq_36:T91743	6.00	81.00	6.5e+03	2500	Lily calcium/calmodulin-dep
N_Geneseq_36:X05858	6.00	82.09	5.7e+03	2154	Human glycine transporter la	N_Geneseq_36:T91743	6.00	80.97	6.5e+03	2524	Staphylococcus aureus conti
N_Geneseq_36:X04721	6.00	82.09	5.7e+03	2164	Bacillus thuringiensis 158C2a	N_Geneseq_36:T91743	6.00	80.97	6.5e+03	2524	Rat NDF clone 20 DNA. New r
N_Geneseq_36:T91797	6.00	82.08	5.7e+03	2164	Full cDNA sequence of the gp30	N_Geneseq_36:Q080227	6.00	80.81	6.7e+03	2531	Vector sequence derived fro
N_Geneseq_36:T91797	6.00	82.07	5.7e+03	2167	Human Ice-ced-3 homologue-1S c	N_Geneseq_36:X19029	6.00	80.80	6.7e+03	2531	DNA -a sequence of Bean Gcl
N_Geneseq_36:X20539	6.00	82.06	5.7e+03	2172	MART-1 melanoma antigen cDNA25	N_Geneseq_36:X19029	6.00	80.76	6.7e+03	2559	Staphylococcus aureus mecB
N_Geneseq_36:T36499	6.00	82.01	5.7e+03	2174	Human interleukin-1 beta conve	N_Geneseq_36:Q04730	6.00	80.72	6.7e+03	2614	Mouse serine protease B5SP-
N_Geneseq_36:X04732	6.00	82.00	5.7e+03	2191	Polynucleotide sequence from t	N_Geneseq_36:Q04730	6.00	80.71	6.7e+03	2615	Sequence encoding glucocort
N_Geneseq_36:T68438	6.00	82.00	5.7e+03	2192	Xenopus orphan receptor 6 cDNA	N_Geneseq_36:T68438	6.00	80.66	6.8e+03	2634	Pseudomonas orfV secretion
N_Geneseq_36:Q06815	6.00	81.99	5.8e+03	2193	Genomic form of lambda-5. Nuci	N_Geneseq_36:T68438	6.00	80.65	6.8e+03	2638	DNA encoding an isopropylma
N_Geneseq_36:T04075	6.00	81.97	5.8e+03	2196	Human gelatinase gene coding reg	N_Geneseq_36:Q22831	6.00	80.62	6.8e+03	2650	Sequence encoding ovine int
N_Geneseq_36:T58559	6.00	81.98	5.8e+03	2196	Sequence encoding heat resista	N_Geneseq_36:Q21003	6.00	80.58	6.9e+03	2652	flg receptor protein gene.
N_Geneseq_36:T43016	6.00	81.98	5.8e+03	2199	M. gallisepticum DNA sequence	N_Geneseq_36:T43016	6.00	80.53	6.9e+03	2682	CDNA encoding a human chron
N_Geneseq_36:X04732	6.00	81.98	5.8e+03	2199	Human breast cancer cell glyco	N_Geneseq_36:T43016	6.00	80.52	6.9e+03	2682	CDNA encoding a human chron
N_Geneseq_36:T68438	6.00	81.98	5.8e+03	2199	Streptococcus pneumoniae polyP	N_Geneseq_36:Q20667	6.00	80.50	6.9e+03	2685	Polynucleotide sequence fro
N_Geneseq_36:T68438	6.00	81.98	5.8e+03	2199	Full cDNA sequence of the gp30	N_Geneseq_36:Q20667	6.00	80.50	6.9e+03	2692	Sequence encoding methylase
N_Geneseq_36:T68438	6.00	81.97	5.8e+03	2201	Hexokinase exon II promoter fr	N_Geneseq_36:X05056	6.00	80.49	6.9e+03	2698	DNA encoding two 3-isopropy
N_Geneseq_36:T68438	6.00	81.97	5.8e+03	2202	Human OVCA1 tumour suppressor	N_Geneseq_36:T68438	6.00	80.47	7.0e+03	2702	Homo sapiens secreted prote
N_Geneseq_36:T68438	6.00	81.92	5.8e+03	2217	Human betaine-GABA transporter	N_Geneseq_36:T10501	6.00	80.44	7.0e+03	2715	P. putida cis/trans isomera
N_Geneseq_36:T68438	6.00	81.91	5.8e+03	2218	Human betaine/GABA transporter	N_Geneseq_36:T10501	6.00	80.39	7.0e+03	2733	Human fibroblast growth fac
N_Geneseq_36:T68438	6.00	81.86	5.9e+03	2235	Canola cold responsive gene BN	N_Geneseq_36:T44351	6.00	80.39	7.0e+03	2733	Human 5-HT2C serotonin rece
N_Geneseq_36:T68438	6.00	81.83	5.9e+03	2245	DNA encoding polypeptide with	N_Geneseq_36:X01067	6.00	80.38	7.0e+03	2735	H. pylori ureA + ureB gene
N_Geneseq_36:T68438	6.00	81.83	5.9e+03	2245	Chromosomal gene fragment of H	N_Geneseq_36:Q80231	6.00	80.36	7.1e+03	2739	Basli methylase gene sequenc
N_Geneseq_36:T68438	6.00	81.82	5.9e+03	2247	Human STCH chaperone protein c	N_Geneseq_36:Q14808	6.00	80.36	7.1e+03	2743	Rat NDF clone 41 DNA. New r
N_Geneseq_36:T68438	6.00	81.79	5.9e+03	2253	Sequence encoding human acetyl	N_Geneseq_36:T108331	6.00	80.36	7.1e+03	2745	B. thuringiensis toxin/gp84
N_Geneseq_36:T68438	6.00	81.79	5.9e+03	2256	Human acetylcholinesterase (A	N_Geneseq_36:Q12485	6.00	80.30	7.1e+03	2767	Glucosylase P gene. Hormoc
N_Geneseq_36:T68438	6.00	81.72	6.0e+03	2278	Enterohaemorrhagic E.coli hlyA	N_Geneseq_36:T67235	6.00	80.26	7.1e+03	2794	DNA encoding A and B subuni
N_Geneseq_36:T68438	6.00	81.70	6.0e+03	2285	Kidney injury associated molec	N_Geneseq_36:T42791	6.00	80.23	7.2e+03	2794	Pasteurella haemolytica tra
N_Geneseq_36:T68438	6.00	81.66	6.0e+03	2297	Human cytokine response gene c	N_Geneseq_36:T51612	6.00	80.16	7.2e+03	2822	Curvularia verruculosa halo
N_Geneseq_36:T68438	6.00	81.66	6.0e+03	2297	Human cytokine response gene c	N_Geneseq_36:T38071	6.00	80.15	7.2e+03	2826	Transferrin binding protein

N_Geneseq_36:Q24234	6.00	80.14	7.3e+03	2830	Lipoxigenase gene. Prodn. of p	N_Geneseq_36:Q28829	6.00	78.57	8.9e+03	3510	Mitochondrial cytochrome P4
N_Geneseq_36:Q24492	6.00	80.14	7.3e+03	2830	Rice lipoxigenase gene. Rice	N_Geneseq_36:T61291	6.00	78.55	8.9e+03	3517	Mouse truncated beta-3 inte
N_Geneseq_36:Q25552	6.00	80.11	7.3e+03	2842	RING4 CDNA clone. DNA encod	N_Geneseq_36:T47554	6.00	78.54	8.9e+03	3521	H6 promoted cytomegalovirus
N_Geneseq_36:T86089	6.00	80.09	7.3e+03	2848	Kex2 protease encoding CDNA. S	N_Geneseq_36:Q06830	6.00	78.54	8.9e+03	3522	Sequence encoding B.thuring
N_Geneseq_36:Q10448	6.00	80.07	7.3e+03	2856	Human basic fibroblast growth	N_Geneseq_36:Q10182	6.00	78.54	8.9e+03	3522	Lepidopteran-active toxin D
N_Geneseq_36:Q14807	6.00	80.03	7.4e+03	2871	B.thuringiensis toxin/gp64 vir	N_Geneseq_36:Q26928	6.00	78.54	8.9e+03	3522	Delta endotoxin gene. Contr
N_Geneseq_36:Q94642	6.00	80.01	7.4e+03	2880	Recombinant cold-resistant pyr	N_Geneseq_36:Q47291	6.00	78.54	8.9e+03	3522	Delta endotoxin gene. Contr
N_Geneseq_36:T12673	6.00	80.01	7.4e+03	2880	Corn derived pyruvate phosphat	N_Geneseq_36:T05250	6.00	78.54	8.9e+03	3522	CryIF/436 chimeric toxin co
N_Geneseq_36:Q57016	6.00	79.98	7.4e+03	2891	PKC delta. Eukaryotic cells th	N_Geneseq_36:T16558	6.00	78.54	8.9e+03	3522	B.t. toxin 81A2 gene. Toxin
N_Geneseq_36:T59145	6.00	79.96	7.4e+03	2900	Thrichia sp. E80 genomic DNA.	N_Geneseq_36:T05270	6.00	78.54	8.9e+03	3522	CryIF/436 chimeric toxin ge
N_Geneseq_36:T32961	6.00	79.96	7.4e+03	2901	Mouse developmental kinase 1 M	N_Geneseq_36:T18702	6.00	78.54	8.9e+03	3522	CryIF/436 chimeric toxin ge
N_Geneseq_36:Q01758	6.00	79.94	7.4e+03	2909	CDNA sequence of rat C kinase	N_Geneseq_36:T18722	6.00	78.54	8.9e+03	3522	Codon-reworked CryIF toxin
N_Geneseq_36:Q80226	6.00	79.92	7.5e+03	2914	Rat NDF clone 19 DNA. New recd	N_Geneseq_36:T45592	6.00	78.54	8.9e+03	3522	Bacillus thuringiensis P881
N_Geneseq_36:Q94641	6.00	79.92	7.5e+03	2915	Recombinant cold-resistant pyr	N_Geneseq_36:T62081	6.00	78.54	8.9e+03	3522	Plasmid pMYC2243 Bt endotox
N_Geneseq_36:T12672	6.00	79.92	7.5e+03	2915	F. bidentis pyruvate phosphat	N_Geneseq_36:T62083	6.00	78.54	8.9e+03	3522	Plasmid pMYC2254 cryIF/436
N_Geneseq_36:T05842	6.00	79.87	7.5e+03	2935	Calcium-independent cytosolic	N_Geneseq_36:T13437	6.00	78.54	8.9e+03	3522	Enterococcus faecalis genom
N_Geneseq_36:T44578	6.00	79.87	7.5e+03	2935	Calcium-independent cytosolic	N_Geneseq_36:Q27658	6.00	78.54	8.9e+03	3525	N-sam cDNA. Gene prod of N-
N_Geneseq_36:T59199	6.00	79.87	7.5e+03	2935	Ca-independent phospholipase A	N_Geneseq_36:Q04284	6.00	78.52	8.9e+03	3531	Gene encoding bacterial tox
N_Geneseq_36:T68827	6.00	79.87	7.5e+03	2935	Hamster cytosolic phospholipase	N_Geneseq_36:Q47290	6.00	78.52	8.9e+03	3531	Delta endotoxin gene. Contr
N_Geneseq_36:T68480	6.00	79.87	7.5e+03	2935	Chinese hamster calcium indepe	N_Geneseq_36:T03929	6.00	78.52	8.9e+03	3531	CryIF(d) gene. New DNA enco
N_Geneseq_36:Q87350	6.00	79.86	7.5e+03	2939	Plasmid BGIN. New intron-modi	N_Geneseq_36:T00293	6.00	78.52	8.9e+03	3531	Bacillus thuringiensis var.
N_Geneseq_36:T37293	6.00	79.86	7.5e+03	2939	pBGINV plasmid used in the cou	N_Geneseq_36:T31159	6.00	78.52	8.9e+03	3531	Bacillus thuringiensis EG11
N_Geneseq_36:T10514	6.00	79.86	7.5e+03	2940	Human IFAK DNA #1. Nucleic ac	N_Geneseq_36:T31160	6.00	78.52	8.9e+03	3531	Bacillus thuringiensis EG11
N_Geneseq_36:Q87790	6.00	79.85	7.5e+03	2945	S. lavendulae mcr locus. Genes	N_Geneseq_36:T31161	6.00	78.52	8.9e+03	3531	Bacillus thuringiensis EG11
N_Geneseq_36:T69280	6.00	79.85	7.5e+03	2945	Human chromodomain Y like (CDY	N_Geneseq_36:T72711	6.00	78.52	8.9e+03	3532	Human inhibitor of apoptosi
N_Geneseq_36:Q65587	6.00	79.77	7.6e+03	2975	Heat resistant aldehyde dehydr	N_Geneseq_36:Q05794	6.00	78.52	8.9e+03	3534	Encodes modified B.t.k. HD-
N_Geneseq_36:T69281	6.00	79.75	7.6e+03	2983	Mouse chromodomain Y like (Cdy	N_Geneseq_36:Q05795	6.00	78.52	8.9e+03	3534	Encodes full-length modifi
N_Geneseq_36:Q04023	6.00	79.74	7.6e+03	2990	Gene pJH10 encoding bacterial	N_Geneseq_36:Q06561	6.00	78.52	8.9e+03	3534	Encodes full-length insecti
N_Geneseq_36:Q95299	6.00	79.71	7.7e+03	3001	Plasmid pAIC2-26 contg. mouse	N_Geneseq_36:T31162	6.00	78.52	8.9e+03	3534	Bacillus thuringiensis EG11
N_Geneseq_36:T16304	6.00	79.71	7.7e+03	3001	Ras coding sequence from AIC2A	N_Geneseq_36:T31163	6.00	78.52	8.9e+03	3534	Bacillus thuringiensis EG11
N_Geneseq_36:Q68842	6.00	79.69	7.7e+03	3007	Plant blight-specific resistanc	N_Geneseq_36:T31165	6.00	78.52	8.9e+03	3534	Bacillus thuringiensis EG11
N_Geneseq_36:T67241	6.00	79.66	7.7e+03	3023	Pasteurella haemolytica transp	N_Geneseq_36:Q56806	6.00	78.51	8.9e+03	3537	Bacillus thuringiensis CryI
N_Geneseq_36:T30093	6.00	79.52	7.9e+03	3079	Human vascular smooth muscle c	N_Geneseq_36:T69681	6.00	78.51	8.9e+03	3537	FltPV spike (S) gene C-termi
N_Geneseq_36:T05628	6.00	79.50	7.9e+03	3088	ADP ribosylation factor gene	N_Geneseq_36:T27594	6.00	78.50	8.9e+03	3540	Homo sapiens pcch-4 coding
N_Geneseq_36:N90387	6.00	79.49	7.9e+03	3091	Gene encoding KEX2 endoproteas	N_Geneseq_36:Q90656	6.00	78.49	9.0e+03	3546	Eph-related PTK Cek10 cDNA.
N_Geneseq_36:T70135	6.00	79.46	7.9e+03	3106	Pyruvate formate lyase homolog	N_Geneseq_36:T61457	6.00	78.49	9.0e+03	3546	C-proteinase clone pCP-2 N
N_Geneseq_36:N90434	6.00	79.43	7.9e+03	3116	Partial sequence of Bacillus t	N_Geneseq_36:Q80648	6.00	78.47	9.0e+03	3558	Gene encoding CryIE/CryIC B
N_Geneseq_36:Q90652	6.00	79.40	8.0e+03	3133	Eph-related tyrosine kinase CD	N_Geneseq_36:Q88414	6.00	78.46	9.0e+03	3561	Human TK2 kinase coding se
N_Geneseq_36:T74654	6.00	79.36	8.0e+03	3150	Staphylococcus aureus contig S	N_Geneseq_36:T02466	6.00	78.45	9.0e+03	3565	Human c-fos oncogene. Antic
N_Geneseq_36:Q80233	6.00	79.33	8.0e+03	3161	Rat NDF clone 42B DNA. New rec	N_Geneseq_36:Q05797	6.00	78.45	9.0e+03	3567	Encodes B.t.entomocidus ins
N_Geneseq_36:T74624	6.00	79.30	8.1e+03	3175	Nucleotide sequence of human K	N_Geneseq_36:Q10181	6.00	78.45	9.0e+03	3567	Lepidopteran-active toxin C
N_Geneseq_36:Q94645	6.00	79.29	8.1e+03	3180	Staphylococcus aureus contig S	N_Geneseq_36:Q47294	6.00	78.45	9.0e+03	3567	Delta endotoxin gene. Contr
N_Geneseq_36:T12676	6.00	79.29	8.1e+03	3180	Recombinant cold-resistant pyr	N_Geneseq_36:Q80646	6.00	78.45	9.0e+03	3567	CryIC gene. New Bacillus th
N_Geneseq_36:Q04092	6.00	79.26	8.1e+03	3191	F. brownii cold resistant pyr	N_Geneseq_36:T27864	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:T62030	6.00	79.11	8.3e+03	3258	3.2 Kb KpnI-PvuII fragment of	N_Geneseq_36:T27846	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:V10525	6.00	79.10	8.3e+03	3264	Human transient receptor poten	N_Geneseq_36:T27847	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:T26587	6.00	79.05	8.3e+03	3287	Human IFAK DNA #2. Nucleic ac	N_Geneseq_36:T27848	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:X23981	6.00	79.02	8.4e+03	3297	Nucleotide sequence of human K	N_Geneseq_36:T27844	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:Q14806	6.00	79.01	8.4e+03	3303	Human HG38 DNA. Human G-protei	N_Geneseq_36:T27843	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:T13441	6.00	78.97	8.4e+03	3319	B.thuringiensis toxin/gp64 vir	N_Geneseq_36:T27863	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:T60295	6.00	78.97	8.4e+03	3319	Interleukin-6 receptor coding	N_Geneseq_36:T27864	6.00	78.45	9.0e+03	3567	DNA encoding a mutated CryI
N_Geneseq_36:Q200914	6.00	78.95	8.4e+03	3328	Human interleukin-6 receptor c	N_Geneseq_36:Q80649	6.00	78.42	9.0e+03	3579	DNA encoding a mutated CryI
N_Geneseq_36:N91089	6.00	78.93	8.5e+03	3337	Human bFGF receptor sequence.	N_Geneseq_36:T31164	6.00	78.42	9.0e+03	3579	Gene encoding CryIA/CryIC B
N_Geneseq_36:Q80228	6.00	78.92	8.5e+03	3344	Fragment of Pseudomonas SMP1	N_Geneseq_36:T31164	6.00	78.40	9.1e+03	3591	Bacillus thuringiensis EG11
N_Geneseq_36:Q80228	6.00	78.87	8.5e+03	3366	Rat NDF clone 22 DNA. New rec	N_Geneseq_36:Q90658	6.00	78.40	9.1e+03	3591	Eph-related PTK Cek10+ cDNA
N_Geneseq_36:Q84803	6.00	78.87	8.5e+03	3366	Spino cerebellar ataxia type 1	N_Geneseq_36:T06481	6.00	78.30	9.2e+03	3639	Cystathionine gamma synthas
N_Geneseq_36:T63191	6.00	78.87	8.5e+03	3367	CDNA from clone dl39-9 which e	N_Geneseq_36:T84934	6.00	78.30	9.2e+03	3642	Human prostate protein HPA3
N_Geneseq_36:T13077	6.00	78.87	8.5e+03	3368	Wheat acetyl-CoA carboxylase A	N_Geneseq_36:T74483	6.00	78.26	9.2e+03	3650	Staphylococcus aureus cont
N_Geneseq_36:T33413	6.00	78.87	8.5e+03	3368	5' end flanking sequence of wh	N_Geneseq_36:T31115	6.00	78.21	9.3e+03	3684	Tobacco necrosis virus stra
N_Geneseq_36:Q62300	6.00	78.84	8.6e+03	3382	pBam-13 plasmid fragment enco	N_Geneseq_36:T14340	6.00	78.18	9.3e+03	3699	Plasmid pBSGFP expressing g
N_Geneseq_36:Q06904	6.00	78.80	8.6e+03	3401	Sequence encoding site specif	N_Geneseq_36:N90339	6.00	78.15	9.4e+03	3718	Sequence of a novel hybrid
N_Geneseq_36:T00771	6.00	78.75	8.7e+03	3423	SGT-sep fusion gene fragment.	N_Geneseq_36:T18977	6.00	78.14	9.4e+03	3722	Non-B. non-C. non-G heparit
N_Geneseq_36:T16273	6.00	78.72	8.7e+03	3437	Pyrodictium occultum 3437 bp D	N_Geneseq_36:T55040	6.00	78.12	9.4e+03	3732	Human HIAP-2 coding sequenc
N_Geneseq_36:Q20545	6.00	78.71	8.7e+03	3442	Encodes KEX2 endopeptidase wit	N_Geneseq_36:T63514	6.00	78.01	9.5e+03	3789	Mycobacterium urease gene c
N_Geneseq_36:Q20545	6.00	78.70	8.7e+03	3447	M.leprae rpoB gene. Rapid dete	N_Geneseq_36:T63514	6.00	77.99	9.5e+03	3800	COCAOAT promoter, pTCCOAOAT
N_Geneseq_36:Q58005	6.00	78.67	8.8e+03	3461	Sequence of plasmid pHN1 show	N_Geneseq_36:Q12760	6.00	77.97	9.6e+03	3809	P40 genomic DNA. Nucleic ac
N_Geneseq_36:Q64068	6.00	78.67	8.8e+03	3461	Non-A. non-B hepatitis virus g	N_Geneseq_36:T07010	6.00	77.91	9.6e+03	3841	ILTV genome fragment. Live
N_Geneseq_36:T30386	6.00	78.67	8.8e+03	3461	5'UTR/CORE/ENV/NS1/NS3/NS3' CD	N_Geneseq_36:T45417	6.00	77.90	9.7e+03	3848	P. suis leukotoxin genes ps
N_Geneseq_36:T30386	6.00	78.65	8.8e+03	3471	Herpes simplex virus type 1 g	N_Geneseq_36:T52593	6.00	77.86	9.7e+03	3868	Homo sapiens pcch-4 gene. N
N_Geneseq_36:N80907	6.00	78.65	8.8e+03	3472	Sequence of Herpes simplex vir	N_Geneseq_36:T12292	6.00	77.85	9.7e+03	3870	Phospholipase C-gamma-1 cDN
N_Geneseq_36:Q84849	6.00	78.65	8.8e+03	3472	Glycoprotein B (gB1). DNA con	N_Geneseq_36:Q46410	6.00	77.84	9.7e+03	3879	Leukotoxin genes hlyB and h
N_Geneseq_36:Q21453	6.00	78.64	8.8e+03	3475	Sequence encoding beta-chain c	N_Geneseq_36:T12293	6.00	77.81	9.8e+03	3893	Phospholipase C-gamma-1 cDN
N_Geneseq_36:T15221	6.00	78.58	8.8e+03	3501	Bacillus thuringiensis 158C2b	N_Geneseq_36:Q10229	6.00	77.75	9.8e+03	3923	Sequence encoding insectici
N_Geneseq_36:T44041	6.00	78.58	8.9e+03	3503	Mouse bFGF receptor DNA. Assay	N_Geneseq_36:Q80523	6.00	77.71	9.9e+03	3946	Human mcl-1 gene. New myelo

N_Geneseq_36:V19872 + 6.00 77.66 1.0e+04 3976 Homo sapiens cdo tumour suppressor
 N_Geneseq_36:Q15153 + 6.00 77.63 1.0e+04 3993 Swine herpes virus-1 major capsid protein
 N_Geneseq_36:V24310 + 6.00 77.62 1.0e+04 3997 Homo sapiens Class II tumour suppressor
 N_Geneseq_36:V26082 + 6.00 77.62 1.0e+04 3997 Tomato pest resistance Mi gene
 N_Geneseq_36:X29857 - 6.00 77.61 1.0e+04 4001 Thermotoga neopolitana type F
 N_Geneseq_36:T09328 - 6.00 77.61 1.0e+04 4003 CrvIE(c) gene. New DNA encoding
 N_Geneseq_36:X13197 - 6.00 77.59 1.0e+04 4013 Enterococcus faecalis genome
 N_Geneseq_36:Q14648 - 6.00 77.58 1.0e+04 4020 CrvIF gene. Purifying and isolating
 N_Geneseq_36:Q52674 - 6.00 77.56 1.0e+04 4029 F-spontin (FP5-9) coding sequence
 N_Geneseq_36:X30095 - 6.00 77.56 1.0e+04 4029 Rat vascular smooth muscle cell
 N_Geneseq_36:N90382 - 6.00 77.54 1.0e+04 4038 Genes encoding human alpha-2-p
 N_Geneseq_36:Q15154 + 6.00 77.53 1.0e+04 4044 Plasmid PHAS2-MCP containing s

seq_name: N_Geneseq_36:V74878

seq_documentation_block:

ID V74878 standard; DNA; 676 BP.
 AC V74878;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #567.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 FH key Location/Qualifiers
 FT misc_feature 421..480
 FT /*tag=a
 FT /note="these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"

EP-786519-A2.

30-JUL-1997.
 07-JAN-1997; 100117.
 05-JAN-1996; US-009861.
 (HUMA-) HUMAN GENOME SCI INC.
 Barash SC, Choi GH, Dillion PJ, Pannon MR, Kunsch CA,
 Rosen CA;
 WPI: 97-374922/35.
 Polynucleotide(s) and proteins derived from Staphylococcus aureus
 stored on computer readable medium and used in the production of
 anti-S.aureus vaccines
 Claim 1; Page 1505-1506; 3271pp; English.
 This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 of the invention. The DNA sequences are recorded on a computer readable
 medium, preferably selected from a floppy or hard disk, random access
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 the S.aureus DNA sequences allows putative functions to be assigned so
 that protein-encoding or regulatory regions of commercial, therapeutic or
 industrial importance can be obtained. Specifically, sequences which are
 likely to encode antigens have been identified and these polypeptides can
 be used in a vaccine composition against S.aureus infection. The
 polypeptides can also be used in a kit for the immunodetection of
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 skin and surgical wound infections, scalded skin syndrome, toxic shock,
 syndrome, etc. Organisms transformed with the DNA sequences can be used
 for recombinant production of the polypeptides. The new DNA sequences
 (and their fragments) are useful as primers or probes for isolating
 homologues of any of the S.aureus DNA sequences contained on the
 computer readable medium.
 SQ Sequence 676 BP; 163 A; 136 C; 77 G; 236 T;

alignment_scores:

Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x V74878/rev

Align seg 1/1 to reverse of: V74878 from: 1 to: 676

176 GluGlyArgAsnArgGlnValArgArgMet 185

|||||
 174 GAAGGTCTGTAATCGCAAGTCAGAGAAATG 145

seq_name: N_Geneseq_36:X20509

seq_documentation_block:

ID X20509 standard; DNA; 12838 BP.
 AC X20509;
 DT 05-MAY-1999 (first entry)
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 OS Treponema pallidum.
 PN WO9859034-A2.
 PD 30-DEC-1998.
 PF 23-JUN-1998; U13041.
 PR 24-JUN-1997; US-050667.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Fraser CM;
 DR WPI: 99-081273/07.
 PT New isolated Treponema pallidum nucleic acids - used to develop
 products for the detection, diagnosis, characterisation, prevention
 and therapy of T. pallidum infections, particularly syphilis
 PS Claim 1; Page 230-237; 1150pp; English.
 CC X20500-21243 represent polynucleotide sequences from the genome of
 Treponema pallidum. The sequences can be used for detection,
 diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 SQ Sequence 12838 BP; 3171 A; 4046 C; 2834 G; 2770 T;

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x X20509

Align seg 1/1 to: X20509 from: 1 to: 12838

55 GlnPheThrAspGluAlaGlyArg 62

|||||
 7113 CAGTTCACCTGACGAGCTGGCCG 7136

seq_name: N_Geneseq_36:Q69432

seq_documentation_block:

ID Q69432 standard; DNA; 46 BP.
 AC Q69432;
 DT 27-FEB-1995 (first entry)
 DE Human heat shock protein (hsp70B) gene, target region.
 KW DNA protein-binding assay; test sequence; screening sequence;
 KW promoter; target; TATA box; Herpes Simplex Virus; HSV;
 KW origin of replication; UL9; transcription factor; TFIID; ds.
 OS Synthetic.
 PN WQ9414980-A.
 PD 07-JUL-1994.
 PF 20-DEC-1993; U12388.
 PR 23-DEC-1992; US-996783.
 PR 17-SEP-1993; US-123936.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
 DR WPI: 94-234711/28.
 PT Sequence-directed DNA-binding molecules - useful in
 PT pharmaceuticals and as molecular reagents
 PS Claim 28; Page 303; 587pp; English.
 CC A DNA protein-binding assay is provided, useful for screening

CC libraries of synthetic or biological cpds. for their ability
 CC to bind DNA test sequences. The assay is versatile in that any
 CC number of test sequences can be tested by placing the test sequence
 CC adjacent to a defined protein-binding screening sequence. Binding
 CC of mols. to these test sequences changes the binding characteristics
 CC of the protein mol. to its cognate binding sequence. When such a mol.
 CC binds the test sequence, the equilibrium of the DNA:protein complexes
 CC is disturbed, generating changes in the concentration of free DNA probe.
 CC One application of this method is to eucaryotic general transcription
 CC factors (e.g. TFIID), where the target region is typically selected
 CC from DNA sequences adjacent to the binding site for the eucaryotic
 CC transcription factor. Numerous exemplary test sequences are given:
 CC the sequences in Q69251-731 and Q69850 correspond to promoter targets
 CC (typically, TATA box-contg. sites) for human genes and the sequences in
 CC Q69732-849 correspond to promoter targets for viral genes. The test
 CC sequences may also be randomly generated. DNA:protein interaction may
 CC be used for screening purposes, e.g. the Herpes Simplex Virus (HSV)
 CC origin of replication and UL9 (see Q69851-52, Q69865 and Q69891).
 SQ Sequence 46 BP; 4 A; 18 C; 20 G; 4 T;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-252-691-7056 x Q69432 ..

Align seg 1/1 to: Q69432 from: 1 to: 46

98 AlaAargLeuthrGlnProGly 104
 |||||
 3 GCCCGGCTGACTCAGCCGGG 23

seq_name: N_Geneseq_36:T63894

seq_documentation_block:
 ID T63894 standard; DNA; 46 BP.
 AC T63894;
 DT 14-MAR-1997 (first entry)
 DE Human hsp70B gene TFIID binding site.
 KW Duplex DNA: target region; binding characteristic; DNA binding protein;
 KW TFIID; transcription factor; binding site; inhibition; enhance; hCG;
 KW cancer; inherited genetic disorder; alpha-D-galactosidase A; ds.
 OS Homo sapiens.
 PN US5578444-A.
 PD 26-NOV-1996.
 PF 27-JUN-1991; 723618.
 PR 23-DEC-1992; US-723618.
 PR 17-SEP-1993; US-996783.
 PR 20-DEC-1993; US-123936.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
 DR WPI: 97-020402/02.
 PT Altering binding characteristics of DNA binding proteins to duplex
 PT DNA - by attaching specific small cpd. to target region close to the
 PT protein's binding site, useful in treatment of viral disease, cancer
 PT etc
 PS Claim 6: Column 191-192; 264pp; English.
 CC The sequences given in T63713-4312 represent duplex DNA's which act
 CC as target regions in the method of the invention. The method for
 CC altering the binding characteristics of a DNA-binding protein to duplex
 CC DNA comprises contacting the duplex DNA with a small molecule which
 CC binds sequence-specifically to a target region, where, when the small
 CC molecule is bound to the target region, it is adjacent to, but not
 CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.
 CC The small molecule is added at a concentration effective to alter the
 CC binding of the DNA binding protein, pref. TFIID, to its binding site on
 CC the duplex DNA. The binding of the small molecule may inhibit or enhance
 CC the binding of the DNA-binding protein to its binding site. The
 CC compounds isolated using this method are potentially useful as ..

CC therapeutic agents for treatment of any disease which involves a
 CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.
 CC The method is suitable for screening large biological or chemical
 CC libraries and allows determination of sequence-specific and relative
 CC affinities of known DNA-binding agents for different DNA sequences.
 CC The design of these duplex DNA's allows a single DNA:protein interaction
 CC to be used for screening sequence-specific, or preferential, DNA binding
 CC proteins that recognise almost any possible sequence (see also T49539-
 CC 74).
 SQ Sequence 46 BP; 4 A; 18 C; 20 G; 4 T;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-252-691-7056 x T63894 ..

Align seg 1/1 to: T63894 from: 1 to: 46

98 AlaAargLeuthrGlnProGly 104
 |||||
 3 GCCCGGCTGACTCAGCCGGG 23

seq_name: N_Geneseq_36:X17182

seq_documentation_block:
 ID X17182 standard; DNA; 46 BP.
 AC X17182;
 DT 06-MAY-1999 (first entry)
 DE Test sequence from human hsp70B gene.
 KW Test sequence; DNA-binding molecule; screening sequence; human;
 KW nucleic acid amplification; target; viral; ds.
 OS Homo sapiens.
 PN US5869241-A.
 PD 09-FEB-1999.
 PF 07-JUN-1995; 475228.
 PR 20-DEC-1993; US-171389.
 PR 27-JUN-1991; US-723618.
 PR 23-DEC-1992; US-996783.
 PR 17-SEP-1993; US-123936.
 PR 07-JUN-1995; US-475228.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
 DR WPI: 99-152755/13.
 PT Determination of DNA sequence preference of a DNA-binding molecule -
 PT based on inhibition of binding of protein to oligonucleotide
 PT sequence attached to test sequence
 PS Claim 3: Columns 193-194; 270pp; English.
 CC Sequences X17001 to X17600 represent specifically claimed target test
 CC sequences that are used in the method of the invention of determining the
 CC DNA sequence preference of a DNA-binding molecule. The method comprises:
 CC (i) adding a test molecule and a DNA-binding protein to a mixture of
 CC duplex DNA test oligonucleotides, each of the test oligonucleotides
 CC having a test sequence adjacent to a screening sequence, where the
 CC screening sequence binds to the DNA-binding protein with a binding
 CC affinity that is independent of the DNA sequence of the test sequence,
 CC and where the mixture of duplex DNA test oligonucleotides includes
 CC several test sequences; (ii) incubating the test molecule, the mixture of
 CC duplex DNA test oligonucleotides and the DNA-binding protein for a time
 CC sufficient to permit binding of the test molecule to test sequences in
 CC the duplex DNA; (iii) separating unbound test oligonucleotides from test
 CC oligonucleotides bound to binding protein; (iv) amplifying the unbound
 CC test oligonucleotides; (v) repeating steps (ii) to (iv); (vi) isolating
 CC the amplified test oligonucleotides; and (vii) sequencing the isolated
 CC test oligonucleotides. Test sequences X17001-X17481 and X17600 correspond
 CC to promoter targets for human genes and test sequences X17482-X17599
 CC correspond to promoter targets for viral genes.
 SQ Sequence 46 BP; 4 A; 18 C; 20 G; 4 T;


```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X17182 ..
Align seg 1/1 to: X17182 from: 1 to: 46
98 AlaArgLeuThrGlnProGly 104
|||||
3 GCCCGGCTGACTAGCCGGG 23
seq_name: N_Geneseq_36:V21143

seq_documentation_block:
ID V21143 standard; DNA: 286 BP.
AC V21143:
DT 30-JUL-1998 (first entry)
DE Human hsp-70B promoter.
KW Heat shock promoter; therapeutic; tumour; ss.
OS Homo sapiens.
PN WO9806864-A2.
PD 19-FEB-1998
PF 14-AUG-1997; U15270.
PR 15-AUG-1996; US-024213.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Moonen C;
DR WPI: 98-168793/15.
PT Spatial and temporal control of construct gene by heat shock
PT promoter - for production of therapeutic protein in cell mass,
PT especially for treating tumours
PS Claim 3; Fig 3c; 58pp; English.
CC The Human Hsp-70B promoter and its analogue from Drosophila (V21142) can
CC be activated using localised heat. The hsp 70B heat shock gene promoter
CC was used to control expression of a gene construct. This can be used for
CC the expression and production of therapeutic proteins in a preselected
CC region of a cell mass. This is especially useful for the targeted
CC expression of therapeutic proteins in a tumour such as a prostate tumour.
SQ Sequence 286 BP; 44 A; 91 C; 109 G; 42 T;

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V21143 ..
Align seg 1/1 to: V21143 from: 1 to: 286
98 AlaArgLeuThrGlnProGly 104
|||||
163 GCCCGGCTGACTAGCCGGG 183
seq_name: N_Geneseq_36:V49571

seq_documentation_block:
ID V49571 standard; cDNA to mRNA; 447 BP.
AC V49571:
DT 21-OCT-1998 (first entry)
DE Human lymphoma cell line U937 clone HP10085 cDNA #2.
KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW haematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
OS Homo sapiens.
PN WO9821328-A2.
PD 22-MAY-1998.
PF 07-NOV-1997; J04056.
PR 13-NOV-1996; JP-301429.

PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR WPI: 98-297932/26.
DR P-PSDB; W64544.
PT Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
PS Claim 3; Page 124; 205pp; English.
CC V49550-V49599 are cDNA sequences which encode human proteins containing a
CC transmembrane domain. These proteins can be used for, e.g. research and
CC nutrition, and may have cytokine and cell proliferation/differentiation,
CC immune stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.
SQ Sequence 447 BP; 161 A; 70 C; 93 G; 123 T;

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V49571 ..
Align seg 1/1 to: V49571 from: 1 to: 447
216 AspValThrProLySLuasn 222
|||||
404 GATGTACACCGAAGAAAT 424
seq_name: N_Geneseq_36:V24918

seq_documentation_block:
ID V24918 standard; DNA; 579 BP.
AC V24918:
DT 30-JUN-1998 (first entry)
DE H. pylori cytoplasmic protein ORF 04gel0816-22086531_f2.10.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW bacterium; ds.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 1..579
FT CD5 /*tag= a
WO9737044-A1.
09-OCT-1997.
27-MAR-1997; U05223.
06-DEC-1996; US-761318.
29-MAR-1996; US-625811.
02-APR-1996; US-758731.
25-OCT-1996; US-736905.
28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI: 97-503122/46.
DR P-PSDB; W5509.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptides - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 5,6,42; Page 327; 1145pp; English.
CC This sequence encodes a H. pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The DNA and probes derived from it may be used for the
CC identification of H. pylori in a sample, and the diagnosis of
CC H. pylori infection. Nucleic acid sequences complementary to the
CC DNA act as antisense sequences, and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can
CC be used in immunoassays to evaluate the abundance and distribution
CC of H. pylori-specific antigens. The genomic sequence of H. pylori

```

CC (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were
 CC analysed for ORF of at least 180 nucleotides, and the predicted
 CC coding regions defined by computer evaluation. To identify likely
 CC H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having
 CC identified and determined the sequences of interest, particular
 CC regions can be isolated from H. pylori by PCR amplification for
 CC recombinant polypeptide production, e.g. in E. coli hosts.
 SQ Sequence 579 BP; 153 A; 112 C; 140 G; 174 T;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-252-691-7056 x V24918 ..

Align seg 1/1 to: V24918 from: 1 to: 579

128 ArgAsnGlyValThrLeuAsn 134
 |||||
 10 AGGAATGCGGTACTTTAAAT 30

seq_name: N_Geneseq_36:T01289

seq_documentation_block:

ID T01289 standard; CDNA; 613 BP.
 AC T01289; 1996 (first entry)
 DT 08-JUL-1996
 DE Rhinicephalus appendiculatus protective antigen IGBP-MC CDNA.
 KW Male tick; protective antigen; metazoan parasite; vaccine; helminth;
 KW arthropod; fly; mite; flea; immunoglobulin binding protein; IGBP;
 KW IGBP-MC; SS.
 OS Rhinicephalus appendiculatus.
 FH Key Location/Qualifiers
 FT cds 1..519
 FT misc_feature 61..120
 FT /tag= a
 FT /tag= b
 FT /note= "identical in N-terminal/translated sequence"

PN W09527056-A1.
 PD 12-OCT-1995.
 PF 04-APR-1995; G00772.
 PR 05-APR-1994; GB-006708.
 PA (MLCW) MALLINKRODT VETERINARY INC.
 PI Nuttall PA, Wang H;
 DR WPI; 95-358631/46.
 DR P-PSDB; R82811.
 PT New metazoan parasite antigens - which bind to host immunoglobulin,
 PT used to prepare vaccines for protection against metazoan parasites
 PS Claim 23; Fig.27C; 93pp; English.
 CC This cDNA encodes IGBP-MC (an immunoglobulin binding protein) from
 CC the male tick. This sequence may be expressed recombinantly for the
 CC production of the protein which is capable of binding host Ig as
 CC a vaccine target, thereby depriving the parasite of an important
 CC means of defence against the immune response of the host. The
 CC encoded antigenic protein can be used in vaccines for stimulating an
 CC immune response against metazoan parasites in humans or animals.
 CC They can be used for protection against helminths and arthropod
 CC parasites e.g. flies, ticks, mites, fleas or bugs.
 SQ Sequence 613 BP; 169 A; 138 C; 166 G; 140 T;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x T01289/rev ..
 Align seg 1/1 to reverse of: T01289 from: 1 to: 613

30 GlnAlaThrArgArgThrPro 36
 |||||
 34 CAAACACACACGGGACACCG 14

seq_name: N_Geneseq_36:V24773

seq_documentation_block:

ID V24773 standard; DNA; 624 BP.
 AC V24773.
 DT 17-JUN-1998 (first entry)
 DE H. pylori ORF 04gel1210orf1.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 FT cds 1..624
 FT /tag= a
 FT /note= "no stop codon given"

W09737044-A1.

09-OCT-1997. U05223.

27-MAR-1997; US-761318.

06-DEC-1996; US-625811.

29-MAR-1996; US-758731.

02-APR-1996; US-736905.

25-OCT-1996; US-738859.

28-OCT-1996; US-738859.

(ASTR) ASTRA AB.

PI Alm RA, Smith D;

DR WPI; 97-503122/46.

P-PSDB; W55364.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection.

PS Claims 5,6; Page 236; 1145pp; English.

CC This sequence encodes a H. pylori protein of unspecified function.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds. The

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC DNA and probes derived from it may be used for the identification of

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

CC acid sequences complementary to the DNA act as antisense sequences and

CC can be used to prevent the translation of H. pylori mRNA. Antibodies

CC against the protein can be used in immunoassays to evaluate the abundance

CC and distribution of H. pylori-specific antigens. The genomic sequence of

CC H. pylori (ATCC 55679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were analysed

CC for ORF of at least 180 nucleotides, and the predicted coding regions

CC defined by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or exported

CC membrane proteins. Having identified and determined the sequences of

CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in E. coli

CC hosts.

SQ Sequence 624 BP; 166 A; 129 C; 148 G; 181 T;

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x V24773 ..

Align seg 1/1 to: V24773 from: 1 to: 624

128 ArgAsnGlyValThrLeuAsn 134

|||||
58 AGGATGGCGTACTTAAAT 78

seq_name: N_Geneseq_36:X21163

seq_documentation_block:

ID X21163 standard; DNA; 641 BP.

AC X21163;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of *Treponema pallidum*.

KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;

enzyme production; ds.

OS *Treponema pallidum*.

PN W09859034-A2.

PD 30-DEC-1998.

PF 23-JUN-1998; U13041.

PR 24-JUN-1997; US-050667.

PA (HUNA-) HUMAN GENOME SCI INC.

PI Fraser CM;

DR WPI; 99-081273/07.

PT New isolated *Treponema pallidum* nucleic acids - used to develop

products for the detection, diagnosis, characterisation, prevention

and therapy of *T. pallidum* infections, particularly syphilis

PS Claim 1; Page 1101; 1150pp; English.

CC X20500-21243 represent polynucleotide sequences from the genome of

Treponema pallidum. The sequences can be used for detection,

diagnosis, characterisation, prevention and therapy for *T. pallidum*

infections, particularly syphilis. They can also be used for detecting

diseases related to *Borrelia* infections in animals, and for the

production of biosynthetic products such as enzymes.

SQ Sequence 641 BP; 151 A; 217 C; 180 G; 92 T;

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x X21163

Align seg 1/1 to: X21163 from: 1 to: 641

35 ThrProGluProGlnProThr 41

|||||

77 ACGCCGAGCCCGACGCGACA 97

seq_name: N_Geneseq_36:X15049

seq_documentation_block:

ID X15049 standard; DNA; 654 BP.

AC X15049;

DT 15-APR-1999 (first entry)

DE DNA encoding a GST-MIG fusion protein.

KW Monokine induced by gamma-interferon; MIG; CXK chemokine; metastasis;

KW angiogenesis inducer; angiostasis inducer; tumour growth inhibition;

KW haemangiomas; rheumatoid arthritis; atherosclerosis; meningioma;

KW idiopathic pulmonary fibrosis; benign prostatic hypertrophy; psoriasis;

KW vascular restenosis; arteriovenous malformation; neovascular glaucoma;

KW angiofibroma; haemophilic joint; hypertrophic scar; Osler-Weber syndrome;

KW pyogenic granuloma retrolental fibroplasia; scleroderma; trachoma;

KW diabetic adhesion; synovitis; dermatitis; endometriosis; pterygium;

KW adult respiratory distress syndrome; chronic bronchitis;

KW cystic fibrosis; fusion protein; ss.

OS Synthetic.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

1. .654

/*tag= a

/note= "sequence contains in-frame insertions"

US5871723-A.

16-FEB-1999.

PF 06-JUN-1995; 468819.

PR 06-JUN-1995; US-468819.

PA (UNMI) UNIV MICHIGAN.

PI Kunkel SL, Polverini PJ, Strieter RM;

DR WPI; 99-166569/14.

DR P-PSDB; W96736.

PT Use of chemokines with a conserved Cys Xaa Cys (CXC) sequence -

which do not contain amino acid sequence ELR, for inhibiting

PT angiogenesis in tumours, rheumatoid arthritis, restenosis or

PT glaucoma

PS Example 13; Columns 127-128; 145pp; English.

CC The present sequence encodes a glutathione-S-transferase (GST)-monokine

CC induced by gamma-interferon (MIG) fusion protein. MIG is a CXC chemokine

CC that is an inhibitor of angiogenesis. The specification describes methods

CC for inhibiting angiogenesis or for inducing angiostasis, using chemokines

CC (with a conserved Cys Xaa Cys (CXC) sequence at the N-terminal) other

CC than platelet factor-4, and which do not contain the amino acid sequence

CC ELR. The methods are useful for inhibiting tumour growth and metastasis

CC and for treating diseases such as haemangiomas, rheumatoid arthritis,

CC atherosclerosis and idiopathic pulmonary fibrosis (IPF), benign prostatic

CC hypertrophy (BPH), vascular restenosis, arteriovenous malformations

CC (AVM), meningioma, neovascular glaucoma, psoriasis, angiofibroma,

CC granuloma retrolental fibroplasia, scleroderma, trachoma, vascular

CC adhesions, synovitis, dermatitis, endometriosis, pterygium, diabetic

CC retinopathy neovascularisation associated with corneal injury or grafts,

CC adult respiratory distress syndrome (ARDS), chronic bronchitis,

CC pseudogout and cystic fibrosis.

CC Sequence 654 BP; 192 A; 135 C; 135 G; 192 T;

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x X15049/rev

Align seg 1/1 to reverse of: X15049 from: 1 to: 654

157 ProProIleArgGluArgLys 163

|||||

196 CCACCTATCAGGACCGAAG 176

seq_name: N_Geneseq_36:X15057

seq_documentation_block:

ID X15057 standard; DNA; 654 BP.

AC X15057;

DT 15-APR-1999 (first entry)

DE DNA encoding GST-MIG fusion protein.

KW monokine induced by gamma-interferon; MIG; CXK chemokine; metastasis;

KW angiogenesis inducer; angiostasis inducer; tumour growth inhibition;

KW haemangiomas; rheumatoid arthritis; atherosclerosis; meningioma;

KW idiopathic pulmonary fibrosis; benign prostatic hypertrophy; psoriasis;

KW vascular restenosis; arteriovenous malformation; neovascular glaucoma;

KW angiofibroma; haemophilic joint; hypertrophic scar; Osler-Weber syndrome;

KW pyogenic granuloma retrolental fibroplasia; scleroderma; trachoma;

KW vascular adhesion; synovitis; dermatitis; endometriosis; pterygium;

KW diabetic retinopathy; neovascularisation; chronic bronchitis;

KW adult respiratory distress syndrome; ARDS; pseudogout;

KW cystic fibrosis; fusion protein; ss.

OS Synthetic.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

1. .627

/*tag= a

/note= "sequence contains numerous in-frame

insertions"

US5871723-A.

16-FEB-1999.

PD 06-JUN-1995; 468819.

CC The DNA sequence can also be used in gene therapy to remove heavy metals
CC from an organ.
SQ Sequence 667 BP; 178 A; 162 C; 134 G; 193 T;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x Q86546 ..

Align seg 1/1 to: Q86546 from: 1 to: 667

136 GlyProThrLeuProAlaGly 142

|||||

447 GGCCCAACTCTGCCAGCTGGT 467

OM of: US-09-252-691-7056 to: Issued_Patents_NA:* out_format : pfs

Date: Jun 10, 2000 2:36 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame+g2n.model -DEV-xlp
-DB-USPTO_Spool/US09252691/runat-05062000_101737_1274/app_query.fasta.1
-DB-Issued_Patents_NA -QPMF-fastap -SUFFIX=oligo.rni
-GAPOPT=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPT=0.000 -GAPOPT=4.500 -QCAPEXT=0.050 -XGAPOPT=60.000
-GAPEXT=60.000 -GAPOPT=6.000 -GAPEXT=7.000 -XGAPOPT=60.000
-GAPEXT=60.000 -DELEX=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=1000 -DOCLIGN=200
-THR_SCORE=quality -THR_MIN=15 -MODE=LOCAL -OUTFT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=100000 -USER=US09252691 -NCPU=6
-ICPU=3 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-252-691-7056

Query length: 222

Database: Issued_Patents_NA.*

Database sequences: 230453

Database length: 64992525

Search time (sec): 26.990000

WARN: XGAPOPT and YGAPOPT must be equal. Assuming YGAPOPT=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/ina/5A/5B/COMB.seq:US-08-171-389-182 +	46	7.00	132.17	6.79	
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-123-936-182 +	46	7.00	132.17	6.79	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-475-228A-182 +	46	7.00	132.17	6.79	
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-482-080A-182 +	46	7.00	132.17	6.79	
/cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:US-08-316-233-56 +	654	7.00	111.56	95.45	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-468-819-73 +	654	7.00	111.56	95.45	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-468-819-75 +	1211	7.00	106.78	176.25	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-23 +	1211	7.00	106.78	176.25	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-25 +	1211	7.00	106.78	176.25	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-27 +	1322	7.00	106.10	192.33	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-241-664B-5 -	1322	7.00	106.10	192.33	
/cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:US-08-943-087-37 +	1322	7.00	106.10	192.33	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-39 +	1468	7.00	105.29	213.47	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-41 +	1631	7.00	104.47	237.05	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-43 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-45 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-47 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-49 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-51 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-53 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-55 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-57 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-59 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-61 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-63 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-65 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-67 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-69 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-71 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-73 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-75 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-77 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-79 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-81 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-83 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-85 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-87 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-89 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-91 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-93 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-95 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-97 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-087-99 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-01 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-03 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-05 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-07 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-09 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-11 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-13 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-15 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-17 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-19 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-21 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-23 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-25 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-27 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-29 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-31 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-33 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-35 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-37 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-39 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-41 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-43 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-45 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-47 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-49 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-51 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-53 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-55 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-57 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-59 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-61 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-63 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-65 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-67 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-69 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-71 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-73 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-75 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-77 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-79 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-81 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-83 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-85 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-87 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-89 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-91 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-93 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-95 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-97 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-943-088-99 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-01 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-03 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-05 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-07 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-09 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-11 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-13 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-15 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-17 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-19 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-21 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-23 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-25 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-27 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-29 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-31 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-33 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-35 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-37 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-39 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-41 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-43 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-45 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-47 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-49 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-51 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-53 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-55 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-57 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-59 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-61 +	1659	7.00	104.34	241.11	
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-944-089-63 +	1659	7.00	104.34	241.11	

[illegible]

[illegible]


```
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-554-612C-51 + 6.00 80.93 4.8e+03 2898 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-09-021-337-9 - 6.00 80.93 4.8e+03 2900 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US96-00419-4 - 6.00 80.92 4.8e+03 2901 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-554-612C-10 + 6.00 80.90 4.8e+03 2909 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-554-612C-11 + 6.00 80.90 4.8e+03 2909 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-550-857A-1 + 6.00 80.89 4.8e+03 2915 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-09-357-073-1 - 6.00 80.86 4.9e+03 2924 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-281-193-1 + 6.00 80.83 4.9e+03 2935 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-422-106-1 + 6.00 80.83 4.9e+03 2935 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-735-716-1 + 6.00 80.83 4.9e+03 2935 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-555-568B-1 + 6.00 80.83 4.9e+03 2935 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US95-08069-1 + 6.00 80.83 4.9e+03 2935 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-119-512-2 + 6.00 80.82 4.9e+03 2939 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-488-015B-2 + 6.00 80.82 4.9e+03 2939 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-07-671-817A-1 + 6.00 80.69 5.0e+03 2990 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-318-826A-7 + 6.00 80.62 5.0e+03 3016 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-370-156-5 + 6.00 80.62 5.0e+03 3016 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-814-095-5 + 6.00 80.62 5.0e+03 3016 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-418-444A-1 + 6.00 80.44 5.1e+03 3088 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-318-826A-6 + 6.00 80.42 5.2e+03 3096 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-370-156-3 + 6.00 80.42 5.2e+03 3096 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-814-095-3 - 6.00 80.42 5.2e+03 3096 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-162-809-1 - 6.00 80.33 5.2e+03 3133 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-693-174-5 + 6.00 80.27 5.3e+03 3157 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-508-857A-9 + 6.00 80.21 5.3e+03 3188 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-729-955A-2 + 6.00 80.02 5.4e+03 3250 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-815-667-14 + 6.00 79.96 5.5e+03 3283 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-611-107-33 + 6.00 79.94 5.5e+03 3292 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-326-670A-1 - 6.00 79.88 5.5e+03 3318 !
/cgn2_6/pdata1/1/na/backfiles1.seq:5457037-1 - 6.00 79.84 5.5e+03 3336 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-469-802B-1 + 6.00 79.77 5.6e+03 3366 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-267-803B-1 + 6.00 79.77 5.6e+03 3366 !
/cgn2_6/pdata1/1/na/backfiles1.seq:5190871-1 - 6.00 79.69 5.7e+03 3401 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-313-185-57 + 6.00 79.58 5.7e+03 3447 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-389-564B-1 + 6.00 79.55 5.8e+03 3461 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-466-047B-1 + 6.00 79.53 5.8e+03 3472 !
/cgn2_6/pdata1/1/na/backfiles1.seq:5244792-1 - 6.00 79.52 5.8e+03 3475 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-07-960-389-1 + 6.00 79.46 5.8e+03 3501 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-448-170-5 - 6.00 79.46 5.8e+03 3503 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-07-631-717A-1 - 6.00 79.46 5.8e+03 3503 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-166-717D-1 - 6.00 79.42 5.8e+03 3521 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-658-665-64 - 6.00 79.42 5.8e+03 3522 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-07-828-788A-9 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-040-751-4 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-349-867-24 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-349-867-28 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-239-476-24 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-239-476-28 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-291-368-1 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-356-034-7 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-598-305A-24 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-598-305A-28 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-639-923A-24 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-639-923A-28 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-962-190-1 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US92-11337-9 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US95-05431-24 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US95-05431-28 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US95-10310-1 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/backfiles1.seq:1516180-3 - 6.00 79.42 5.9e+03 3522 !
/cgn2_6/pdata1/1/na/backfiles1.seq:US-07-828-788A-7 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-07-920-085-1 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-430-493-3 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-841-178-27 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-754-490-9 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-08-754-490-11 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-08-754-490-13 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US92-11337-7 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/backfiles1.seq:1516929-1 - 6.00 79.40 5.9e+03 3531 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-09-205-204-1 - 6.00 79.40 5.9e+03 3532 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-476-062A-40 - 6.00 79.39 5.9e+03 3533 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US96-01314-40 - 6.00 79.39 5.9e+03 3533 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-841-178-24 - 6.00 79.39 5.9e+03 3534 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-841-178-26 - 6.00 79.39 5.9e+03 3534 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-841-178-28 - 6.00 79.39 5.9e+03 3534 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-08-754-490-25 - 6.00 79.39 5.9e+03 3534 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-08-754-490-27 - 6.00 79.39 5.9e+03 3534 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-377-690-3 - 6.00 79.38 5.9e+03 3537 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-162-809-38 - 6.00 79.38 5.9e+03 3537 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-162-809-9 - 6.00 79.36 5.9e+03 3546 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-602-737-5 - 6.00 79.34 5.9e+03 3558 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-097-937A-12 + 6.00 79.33 5.9e+03 3561 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-306-691B-44 + 6.00 79.32 5.9e+03 3565 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US93-06251-20 + 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-07-828-788A-15 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-356-034-5 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-602-737-1 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-1 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-3 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-5 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-7 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-9 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-11 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-58 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-980-071-60 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-757-536-1 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-757-536-5 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-757-536-7 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-757-536-9 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-757-536-11 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-1 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-3 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-5 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-7 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-9 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-11 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-58 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-314-093-60 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US92-11337-15 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/backfiles1.seq:5188960-5 - 6.00 79.32 5.9e+03 3567 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-602-737-7 - 6.00 79.29 5.9e+03 3579 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-754-490-29 - 6.00 79.29 5.9e+03 3579 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-162-809-13 - 6.00 79.27 6.0e+03 3591 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-07-951-715A-6 - 6.00 79.20 6.0e+03 3624 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-459-448A-6 - 6.00 79.20 6.0e+03 3624 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-459-595A-6 - 6.00 79.20 6.0e+03 3624 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-737-524B-26 + 6.00 79.16 6.0e+03 3639 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-08-946-026-16 - 6.00 79.16 6.1e+03 3642 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-08-646-538-6 - 6.00 79.04 6.1e+03 3599 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-680-326-29 + 6.00 79.02 6.2e+03 3708 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-215-805A-1 + 6.00 78.73 6.4e+03 3848 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-138-641-1 + 6.00 78.69 6.4e+03 3870 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-138-641-3 + 6.00 78.69 6.4e+03 3870 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-138-641-5 + 6.00 78.64 6.5e+03 3893 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-138-641-3 + 6.00 78.64 6.5e+03 3893 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-465-609-4 - 6.00 78.58 6.5e+03 3923 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-09-226-568-18 - 6.00 78.56 6.5e+03 3934 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-077-848A-1 - 6.00 78.54 6.6e+03 3946 !
/cgn2_6/pdata1/1/na/6_COMB.seq:US-09-211-640-1 - 6.00 78.54 6.6e+03 3946 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US94-03547-1 - 6.00 78.42 6.6e+03 4003 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-257-999-1 - 6.00 78.39 6.7e+03 4039 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:PCT-US91-02560-1 - 6.00 78.37 6.7e+03 4039 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-07-862-031B-9 - 6.00 78.37 6.7e+03 4039 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-313-288B-9 - 6.00 78.37 6.7e+03 4039 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US93-03164-9 - 6.00 78.37 6.7e+03 4039 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US91-09422-18 - 6.00 78.25 6.8e+03 4106 !
/cgn2_6/pdata1/1/na/5B_COMB.seq:US-08-434-823-1 - 6.00 78.23 6.8e+03 4106 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-457-366-1 - 6.00 78.23 6.8e+03 4106 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-702-572-14 - 6.00 78.12 6.9e+03 4164 !
/cgn2_6/pdata1/1/na/5A_COMB.seq:US-08-204-675-1 - 6.00 78.12 6.9e+03 4164 !
/cgn2_6/pdata1/1/na/5C_COMB.seq:US-08-660-754-1 - 6.00 78.12 6.9e+03 4164 !
/cgn2_6/pdata1/1/na/5D_COMB.seq:US-08-796-364-1 - 6.00 78.12 6.9e+03 4164 !
/cgn2_6/pdata1/1/na/PCTUS_COMB.seq:PCT-US95-02520-1 - 6.00 78.12 6.9e+03 4164 !
```


/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-944-968A-11 + 5.00 99.06 474.31 24 1 S
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-944-423A-11 + 5.00 99.06 474.31 24 1 S
/cgn2_6/ptodata/1/ina/PTUS_COMB.seq:PCT-US94-10957-1 - 5.00 99.06 474.31 24 1 S
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-182-961B-17 + 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-623-891B-64 - 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-692-787-52 - 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-713-557B-29 - 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-668-650B-18 - 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-859-998-329 + 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-859-998-799 + 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-922-955-1 + 5.00 98.74 493.98 25 1 S
/cgn2_6/ptodata/1/ina/PTUS_COMB.seq:PCT-US94-07659-19 + 5.00 98.74 493.98 25 1 S

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-171-389-182

seq_documentation_block: Sequence 182, Application US/08171389
; Patent No. 5578444
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,389
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human hsp70B gene
; US-08-171-389-182

alignment_scores: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-252-691-7056 x US-08-171-389-182 ..
Align seg 1/1 to: US-08-171-389-182 from: 1 to: 46
98 AlaargLeuThrGlnProGly 104
|||||
3 GCCCGGCTGACTCAGCCGGG 23
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-123-936-182
seq_documentation_block: Sequence 182, Application US/08123936
; Patent No. 5726014
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; TITLE OF INVENTION: Screening Assay for the Detection of
; TITLE OF INVENTION: DNA-Binding Molecules
; NUMBER OF SEQUENCES: 640
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,936
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human hsp70B gene
; US-08-123-936-182
alignment_scores: 7.00 Length: 7
Quality: 7.00

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 11:25:46 ; Search time 12.47 Seconds
(without alignments)
257.025 Million cell updates/sec

Title: US-09-252-691-7056
Perfect score: 222
Sequence: 1 A1MRQLTPENTWTKTSFRK.....SYTILDSLANGWRDVTPEKN 222

Scoring table: OIIGO
Gapop 60.0 , Gapext 60.0

Searched: 145308 seqs, 14437401 residues

Word size : 0

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/PCUS_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	3.2	416	2	US-08-225-488-7
2	6	2.7	22	1	US-08-769-309A-21
3	6	2.7	99	1	US-08-449-639-3
4	6	2.7	93	2	US-08-902-623-43
5	6	2.7	165	1	US-08-215-805A-82
6	6	2.7	179	2	US-08-933-750C-28
7	6	2.7	242	2	US-08-622-352A-3
8	6	2.7	292	1	US-08-769-309A-3
9	6	2.7	327	1	US-08-375-962B-12
10	6	2.7	327	2	US-08-562-114B-12
11	6	2.7	339	2	US-08-855-714-3
12	6	2.7	365	2	US-08-467-559B-9
13	6	2.7	403	2	US-08-846-762-89
14	6	2.7	409	2	US-09-031-485-15
15	6	2.7	409	3	US-08-847-429A-15
16	6	2.7	446	3	US-08-956-254-2
17	6	2.7	448	2	US-09-015-815-1
18	6	2.7	500	2	US-08-987-519-2
19	6	2.7	500	2	US-08-987-519-2
20	6	2.7	501	2	US-08-987-519-3
21	6	2.7	507	2	US-08-409-122-2
22	6	2.7	507	2	US-08-408-669-2
23	6	2.7	516	2	US-09-019-201A-3
24	6	2.7	532	3	US-08-737-336-6
25	6	2.7	586	2	US-08-630-822A-70
26	6	2.7	586	2	US-09-005-069-70
27	6	2.7	597	1	US-08-374-155A-12
28	6	2.7	597	2	US-08-785-396-12

Sequence 10, Appl	1	US-08-374-155A-10	600	2.7
Sequence 10, Appl	2	US-08-785-396-10	600	2.7
Sequence 2, Appl	614	US-08-543-881-2	614	2.7
Sequence 2, Appl	614	US-08-291-289-2	614	2.7
Sequence 2, Appl	614	PCT-US94-00119-2	614	2.7
Sequence 2, Appl	614	PCT-US95-10579-2	614	2.7
Sequence 4, Appl	629	US-08-374-155A-4	629	2.7
Sequence 4, Appl	629	US-08-785-396-4	629	2.7
Sequence 27, Appl	661	US-08-417-174-27	661	2.7
Sequence 121, Appl	661	US-08-417-174-121	661	2.7
Sequence 2, Appl	661	US-08-525-742-2	661	2.7
Sequence 27, Appl	661	US-08-231-565A-27	661	2.7
Sequence 6, Appl	668	US-09-007-961-27	668	2.7
Sequence 1, Appl	879	US-07-891-942G-6	879	2.7
Sequence 2, Appl	897	US-08-554-612C-1	897	2.7
Sequence 4, Appl	947	US-07-960-389-2	947	2.7
Sequence 2, Appl	953	US-08-500-857A-4	953	2.7
Sequence 2, Appl	955	US-08-500-857A-10	955	2.7
Sequence 10, Appl	980	US-08-220-151-5	980	2.7
Sequence 5, Appl	980	US-08-413-118-5	980	2.7
Sequence 5, Appl	1148	US-08-313-185-58	1148	2.7
Sequence 2, Appl	1243	US-08-557-139-2	1243	2.7
Sequence 2, Appl	1290	US-08-138-641-2	1290	2.7
Sequence 9, Appl	1290	US-08-138-133-2	1290	2.7
Sequence 33, Appl	1745	US-09-031-485-33	1745	2.7
Sequence 33, Appl	1745	US-08-847-429A-33	1745	2.7
Sequence 5, Appl	1780	US-08-769-309A-5	1780	2.7
Sequence 2, Appl	2206	US-07-852-280-2	2206	2.7
Sequence 3, Appl	2938	US-08-461-503-2	2938	2.7
Sequence 1, Appl	3011	PCT-US94-00198-3	3011	2.7
Sequence 1, Appl	3011	US-08-188-281B-1	3011	2.7
Sequence 1, Appl	3011	US-08-453-552-1	3011	2.7
Sequence 1, Appl	3011	US-08-710-637-1	3011	2.7
Sequence 1, Appl	3011	PCT-US93-00907-1	3011	2.7
Sequence 1, Appl	3011	PCT-US94-07280-1	3011	2.7
Sequence 1, Appl	3011	PCT-US95-01087-1	3011	2.7
Sequence 6, Appl	3144	US-08-246-982A-6	3144	2.7
Sequence 6, Appl	3144	US-08-453-265-6	3144	2.7
Sequence 42, Appl	3144	US-08-457-273B-42	3144	2.7
Sequence 4, Appl	3567	US-07-642-734C-4	3567	2.7
Sequence 18, Appl	7	US-08-067-387-18	7	2.3
Sequence 18, Appl	7	PCT-US94-05796-18	7	2.3
Sequence 6, Appl	9	US-09-070-756-6	9	2.3
Sequence 170, Appl	10	US-08-556-597-170	10	2.3
Sequence 73, Appl	10	US-08-751-767A-73	10	2.3
Sequence 12, Appl	11	US-08-156-552A-12	11	2.3
Sequence 36, Appl	12	US-08-260-582-36	12	2.3
Sequence 189, Appl	12	PCT-US94-05471-36	12	2.3
Sequence 36, Appl	13	US-08-764-640-173	13	2.3
Sequence 173, Appl	13	US-08-764-640-173	13	2.3
Sequence 179, Appl	13	US-08-764-640-179	13	2.3
Sequence 180, Appl	13	US-08-764-640-180	13	2.3
Sequence 183, Appl	13	US-08-764-640-183	13	2.3
Sequence 228, Appl	13	US-08-764-640-228	13	2.3
Sequence 230, Appl	13	US-08-764-640-230	13	2.3
Sequence 6, Appl	14	US-08-463-620-6	14	2.3
Sequence 7, Appl	14	US-08-463-620-7	14	2.3
Sequence 12, Appl	14	US-08-764-640-12	14	2.3
Sequence 38, Appl	14	US-08-764-640-38	14	2.3
Sequence 52, Appl	14	US-08-764-640-52	14	2.3
Sequence 172, Appl	14	US-08-764-640-172	14	2.3
Sequence 174, Appl	14	US-08-764-640-174	14	2.3
Sequence 175, Appl	14	US-08-764-640-175	14	2.3
Sequence 176, Appl	14	US-08-764-640-176	14	2.3
Sequence 177, Appl	14	US-08-764-640-177	14	2.3
Sequence 178, Appl	14	US-08-764-640-178	14	2.3
Sequence 181, Appl	14	US-08-764-640-181	14	2.3
Sequence 182, Appl	14	US-08-764-640-182	14	2.3
Sequence 184, Appl	14	US-08-764-640-184	14	2.3

102	5	2.3	14	2	US-08-764-640-229	Sequence 229, App	175	5	2.3	33	1	US-08-460-602A-81	Sequence 81, Appl
103	5	2.3	14	2	US-08-934-222-23	Sequence 23, Appl	176	5	2.3	33	1	US-08-460-602A-83	Sequence 83, Appl
104	5	2.3	14	2	US-08-933-402-23	Sequence 23, Appl	177	5	2.3	33	1	US-08-460-602A-84	Sequence 84, Appl
105	5	2.3	14	2	US-09-207-621-23	Sequence 23, Appl	178	5	2.3	33	1	US-08-463-966A-41	Sequence 41, Appl
106	5	2.3	14	2	US-08-224-917-6	Sequence 6, Appl	179	5	2.3	33	1	US-08-463-966A-80	Sequence 80, Appl
107	5	2.3	14	2	US-08-224-917-7	Sequence 7, Appl	180	5	2.3	33	1	US-08-463-966A-81	Sequence 81, Appl
108	5	2.3	14	2	US-08-533-818-23	Sequence 7, Appl	181	5	2.3	33	1	US-08-463-966A-83	Sequence 83, Appl
109	5	2.3	14	2	US-08-914-853-6	Sequence 6, Appl	182	5	2.3	33	1	US-08-463-966A-84	Sequence 84, Appl
110	5	2.3	14	2	US-08-914-853-7	Sequence 7, Appl	183	5	2.3	33	1	US-08-465-217A-41	Sequence 41, Appl
111	5	2.3	14	2	US-08-914-853-7	Sequence 7, Appl	184	5	2.3	33	1	US-08-465-217A-80	Sequence 80, Appl
112	5	2.3	14	4	PCT-US95-03934A-6	Sequence 6, Appl	185	5	2.3	33	1	US-08-465-217A-81	Sequence 81, Appl
113	5	2.3	14	4	PCT-US95-03934A-7	Sequence 7, Appl	186	5	2.3	33	1	US-08-465-217A-83	Sequence 83, Appl
114	5	2.3	15	1	US-08-268-251-17	Sequence 17, Appl	187	5	2.3	33	1	US-08-465-217A-84	Sequence 84, Appl
115	5	2.3	15	2	US-08-637-759B-95	Sequence 95, Appl	188	5	2.3	33	1	US-08-464-329A-41	Sequence 41, Appl
116	5	2.3	15	2	US-09-049-577-5	Sequence 95, Appl	189	5	2.3	33	2	US-08-464-329A-80	Sequence 80, Appl
117	5	2.3	15	4	US-08-871-355A-95	Sequence 17, Appl	190	5	2.3	33	2	US-08-464-329A-81	Sequence 81, Appl
118	5	2.3	15	5	PCT-US93-01112-17	Sequence 17, Appl	191	5	2.3	33	2	US-08-464-329A-83	Sequence 83, Appl
119	5	2.3	16	1	US-08-291-349A-3	Sequence 3, Appl	192	5	2.3	33	2	US-08-464-329A-84	Sequence 84, Appl
120	5	2.3	16	2	US-09-025-706-8	Sequence 8, Appl	193	5	2.3	33	2	US-08-462-507A-41	Sequence 41, Appl
121	5	2.3	16	3	US-08-394-748A-11	Sequence 11, Appl	194	5	2.3	33	2	US-08-462-507A-80	Sequence 80, Appl
122	5	2.3	16	4	PCT-US95-02478-11	Sequence 11, Appl	195	5	2.3	33	2	US-08-462-507A-81	Sequence 81, Appl
123	5	2.3	17	2	US-08-480-190-123	Sequence 123, App	196	5	2.3	33	2	US-08-462-507A-83	Sequence 83, Appl
124	5	2.3	17	2	US-08-488-379-123	Sequence 123, App	197	5	2.3	33	2	US-08-462-507A-84	Sequence 84, Appl
125	5	2.3	17	4	PCT-US93-07545-123	Sequence 123, App	198	5	2.3	33	2	US-08-467-881A-81	Sequence 81, Appl
126	5	2.3	18	1	US-08-257-528B-50	Sequence 50, Appl	199	5	2.3	33	2	US-08-467-881A-83	Sequence 83, Appl
127	5	2.3	18	1	US-08-460-602A-50	Sequence 50, Appl	200	5	2.3	33	2	US-08-467-881A-84	Sequence 84, Appl
128	5	2.3	18	1	US-08-463-966A-50	Sequence 50, Appl	201	5	2.3	33	2	US-08-467-881A-88	Sequence 88, Appl
129	5	2.3	18	1	US-08-465-217A-50	Sequence 50, Appl	202	5	2.3	33	2	US-08-467-881A-89	Sequence 89, Appl
130	5	2.3	18	2	US-08-764-640-6	Sequence 6, Appl	203	5	2.3	34	1	US-08-118-270-184	Sequence 184, App
131	5	2.3	18	2	US-08-764-640-7	Sequence 7, Appl	204	5	2.3	34	1	US-08-257-528B-45	Sequence 45, Appl
132	5	2.3	18	2	US-08-764-640-44	Sequence 44, Appl	205	5	2.3	34	1	US-08-460-602A-45	Sequence 45, Appl
133	5	2.3	18	2	US-08-764-640-53	Sequence 53, Appl	206	5	2.3	34	1	US-08-463-966A-45	Sequence 45, Appl
134	5	2.3	18	2	US-08-764-640-58	Sequence 58, Appl	207	5	2.3	34	1	US-08-465-217A-45	Sequence 45, Appl
135	5	2.3	18	2	US-08-764-640-58	Sequence 58, Appl	208	5	2.3	34	2	US-08-464-329A-45	Sequence 45, Appl
136	5	2.3	18	2	US-08-482-507A-50	Sequence 50, Appl	209	5	2.3	34	2	US-08-462-507A-45	Sequence 45, Appl
137	5	2.3	18	2	US-08-764-640-7	Sequence 7, Appl	210	5	2.3	34	2	US-08-462-507A-45	Sequence 45, Appl
138	5	2.3	19	2	US-08-764-640-8	Sequence 8, Appl	211	5	2.3	34	2	US-08-467-881A-45	Sequence 45, Appl
139	5	2.3	19	2	US-08-764-640-48	Sequence 48, Appl	212	5	2.3	36	1	PCT-US93-08528-184	Sequence 184, App
140	5	2.3	19	2	US-08-764-640-56	Sequence 56, Appl	213	5	2.3	36	1	US-08-781-020-12	Sequence 12, Appl
141	5	2.3	19	2	US-08-764-640-164	Sequence 164, App	214	5	2.3	37	1	US-08-213-236-10	Sequence 10, Appl
142	5	2.3	19	3	US-08-938-830-21	Sequence 21, App	215	5	2.3	43	2	US-08-533-996A-2	Sequence 2, Appl
143	5	2.3	20	1	US-07-678-374D-23	Sequence 23, App	216	5	2.3	43	2	US-08-389-360-2	Sequence 2, Appl
144	5	2.3	20	2	US-08-726-306A-31	Sequence 31, App	217	5	2.3	45	1	US-08-058-200-102	Sequence 102, App
145	5	2.3	20	2	US-08-945-168-28	Sequence 28, App	218	5	2.3	45	2	US-08-770-544-14	Sequence 14, Appl
146	5	2.3	21	2	US-08-615-942A-7	Sequence 7, Appl	219	5	2.3	45	2	US-08-800-644-102	Sequence 102, App
147	5	2.3	21	2	US-08-726-306A-15	Sequence 15, Appl	220	5	2.3	48	2	US-08-460-890A-49	Sequence 49, Appl
148	5	2.3	22	1	US-07-828-450-15	Sequence 15, Appl	221	5	2.3	48	3	US-08-167-641C-49	Sequence 49, Appl
149	5	2.3	22	1	US-08-396-630A-8	Sequence 8, Appl	222	5	2.3	49	2	US-08-460-890A-48	Sequence 48, Appl
150	5	2.3	23	1	US-08-176-500-89	Sequence 89, Appl	223	5	2.3	49	3	US-08-167-641C-48	Sequence 48, Appl
151	5	2.3	23	1	US-08-268-251-47	Sequence 47, Appl	224	5	2.3	52	1	PCT-US92-07813-6	Sequence 6, Appl
152	5	2.3	23	1	US-08-471-052A-89	Sequence 89, Appl	225	5	2.3	52	1	US-08-381-666-1	Sequence 1, Appl
153	5	2.3	23	1	US-08-189-331-89	Sequence 89, Appl	226	5	2.3	52	1	US-08-309-420-7	Sequence 7, Appl
154	5	2.3	23	1	US-08-471-939-89	Sequence 89, Appl	227	5	2.3	52	4	US-08-309-419-7	Sequence 7, Appl
155	5	2.3	23	2	US-08-471-800-89	Sequence 89, Appl	228	5	2.3	52	4	PCT-US95-11856-7	Sequence 7, Appl
156	5	2.3	23	2	US-08-471-068-89	Sequence 89, Appl	229	5	2.3	53	1	PCT-US95-11878-7	Sequence 7, Appl
157	5	2.3	23	4	PCT-US93-01112-47	Sequence 47, App	230	5	2.3	53	1	US-07-857-060-1	Sequence 1, Appl
158	5	2.3	25	1	US-08-764-640-231	Sequence 231, App	231	5	2.3	53	2	US-08-726-306A-144	Sequence 144, App
159	5	2.3	26	1	US-08-453-289-1	Sequence 258, App	232	5	2.3	53	3	US-08-630-916A-79	Sequence 79, Appl
160	5	2.3	26	1	US-07-942-245-258	Sequence 258, App	233	5	2.3	54	1	US-08-381-370-1	Sequence 1, Appl
161	5	2.3	26	2	US-08-574-086-1	Sequence 1, Appl	234	5	2.3	54	1	US-08-262-037-54	Sequence 54, Appl
162	5	2.3	26	2	US-08-559-524A-13	Sequence 13, Appl	235	5	2.3	54	2	US-08-456-647B-38	Sequence 38, Appl
163	5	2.3	27	1	US-07-668-648-12	Sequence 12, App	236	5	2.3	54	2	US-08-237-401A-38	Sequence 38, Appl
164	5	2.3	27	2	US-08-429-998-12	Sequence 12, App	237	5	2.3	58	2	US-08-691-814B-28	Sequence 28, Appl
165	5	2.3	27	2	US-08-431-333-12	Sequence 12, App	238	5	2.3	62	1	US-08-381-041A-1	Sequence 1, Appl
166	5	2.3	27	4	PCT-US91-02321-12	Sequence 12, App	239	5	2.3	63	1	US-08-381-049-1	Sequence 1, Appl
167	5	2.3	32	2	US-08-606-143-38	Sequence 38, App	240	5	2.3	63	1	US-08-257-528B-91	Sequence 91, Appl
168	5	2.3	33	1	US-08-257-528B-40	Sequence 41, App	241	5	2.3	63	1	US-08-460-602A-91	Sequence 91, Appl
169	5	2.3	33	1	US-08-257-528B-81	Sequence 81, App	242	5	2.3	63	1	US-08-463-966A-91	Sequence 91, Appl
170	5	2.3	33	1	US-08-257-528B-83	Sequence 83, App	243	5	2.3	63	2	US-08-465-217A-91	Sequence 91, Appl
171	5	2.3	33	1	US-08-257-528B-84	Sequence 84, App	244	5	2.3	63	2	US-08-464-329A-91	Sequence 91, Appl
172	5	2.3	33	1	US-08-257-528B-84	Sequence 84, App	245	5	2.3	63	2	US-08-462-507A-91	Sequence 91, Appl
173	5	2.3	33	1	US-08-460-602A-41	Sequence 41, App	246	5	2.3	64	1	US-08-467-881A-91	Sequence 91, Appl
174	5	2.3	33	1	US-08-460-602A-80	Sequence 80, App	247	5	2.3	64	1	US-08-257-528B-92	Sequence 92, Appl

248	1	US-08-463-966A-92	Sequence 92, Appl	321	5	2.3	119	2	US-08-347-563A-25	Sequence 25, Appl
249	1	US-08-465-217A-92	Sequence 92, Appl	322	5	2.3	120	2	US-08-232-087A-8	Sequence 8, Appl
250	2	US-08-464-329A-92	Sequence 92, Appl	323	5	2.3	121	1	US-08-350-884-72	Sequence 72, Appl
251	64	US-08-462-507A-92	Sequence 92, Appl	324	5	2.3	122	1	US-08-266-451B-18	Sequence 18, Appl
252	64	US-08-467-881A-92	Sequence 92, Appl	325	5	2.3	122	1	US-08-497-025-5	Sequence 5, Appl
253	66	US-08-262-037-50	Sequence 50, Appl	326	5	2.3	122	1	US-08-709-173-72	Sequence 72, Appl
254	68	US-08-330-638D-2	Sequence 2, Appl	327	5	2.3	122	1	US-08-748-725-18	Sequence 18, Appl
255	68	US-08-606-789-8	Sequence 8, Appl	328	5	2.3	122	2	US-08-709-177-72	Sequence 72, Appl
256	68	US-09-111-348-8	Sequence 8, Appl	329	5	2.3	122	5	5169835-12	Patent No. 5169835
257	68	US-08-511-485-28	Sequence 28, Appl	330	5	2.3	125	1	US-08-478-039-69	Sequence 69, Appl
258	68	US-08-906-746A-2	Sequence 2, Appl	331	5	2.3	125	1	US-08-476-349A-69	Sequence 69, Appl
259	69	US-08-350-884-74	Sequence 74, Appl	332	5	2.3	126	2	US-08-383-649-1	Sequence 1, Appl
260	69	US-08-709-173-74	Sequence 74, Appl	333	5	2.3	127	2	US-08-853-659A-56	Sequence 56, Appl
261	69	US-08-709-177-74	Sequence 74, Appl	334	5	2.3	128	1	US-08-384-183-1	Sequence 1, Appl
262	69	PCT-US95-06406A-6	Sequence 6, Appl	335	5	2.3	131	1	US-08-384-492-1	Sequence 1, Appl
263	70	US-08-315-695-3	Sequence 3, Appl	336	5	2.3	131	2	US-08-394-189B-21	Sequence 21, Appl
264	72	US-08-280-443-11	Sequence 11, Appl	337	5	2.3	132	1	US-08-392-419-4	Sequence 4, Appl
265	72	US-08-457-459-11	Sequence 11, Appl	338	5	2.3	133	1	US-08-253-877C-10	Sequence 10, Appl
266	72	US-08-555-678-11	Sequence 11, Appl	339	5	2.3	133	2	US-08-452-164A-10	Sequence 10, Appl
267	72	PCT-US95-02275-11	Sequence 11, Appl	340	5	2.3	133	3	US-08-603-024-4	Sequence 4, Appl
268	74	US-08-381-037-1	Sequence 1, Appl	341	5	2.3	135	1	US-08-683-650-1	Sequence 1, Appl
269	74	PCT-US92-00282-24	Sequence 24, Appl	342	5	2.3	136	1	US-08-426-627-17	Sequence 17, Appl
270	75	US-08-381-054-1	Sequence 1, Appl	343	5	2.3	136	2	US-08-580-545B-8	Sequence 8, Appl
271	76	US-07-881-075-6	Sequence 6, Appl	344	5	2.3	138	1	US-08-384-292-1	Sequence 1, Appl
272	76	US-08-120-827-6	Sequence 6, Appl	345	5	2.3	138	1	US-08-350-884-78	Sequence 78, Appl
273	76	US-08-478-675-6	Sequence 6, Appl	346	5	2.3	138	1	US-08-468-709B-2	Sequence 2, Appl
274	81	US-08-469-667-14	Sequence 14, Appl	347	5	2.3	138	1	US-08-468-709B-12	Sequence 12, Appl
275	81	US-09-014-969-32	Sequence 32, Appl	348	5	2.3	138	1	US-08-709-173-78	Sequence 78, Appl
276	81	PCT-US95-07289-14	Sequence 14, Appl	349	5	2.3	138	2	US-08-241-664B-2	Sequence 2, Appl
277	83	US-08-665-259-14	Sequence 14, Appl	350	5	2.3	138	2	US-08-241-664B-12	Sequence 12, Appl
278	84	US-08-901-306-4	Sequence 4, Appl	351	5	2.3	138	4	US-08-709-177-78	Sequence 78, Appl
279	85	US-08-381-163-1	Sequence 1, Appl	352	5	2.3	138	4	PCT-US93-03936-2	Sequence 2, Appl
280	88	5206352-2	Patent No. 5206352	353	5	2.3	139	2	US-08-805-918-2	Sequence 2, Appl
281	90	US-08-718-538-7	Sequence 7, Appl	354	5	2.3	140	3	US-09-053-197A-34	Sequence 34, Appl
282	91	US-08-633-682-5	Sequence 5, Appl	355	5	2.3	141	1	US-08-438-123-7	Sequence 7, Appl
283	91	US-08-936-772-5	Sequence 5, Appl	356	5	2.3	141	1	US-08-383-632-1	Sequence 1, Appl
284	93	US-08-381-040-1	Sequence 1, Appl	357	5	2.3	143	5	5169835-18	Patent No. 5169835
285	93	US-08-350-884-76	Sequence 76, Appl	358	5	2.3	144	1	US-08-442-916-1	Sequence 1, Appl
286	93	US-08-709-173-76	Sequence 76, Appl	359	5	2.3	144	2	US-08-905-694-1	Sequence 1, Appl
287	93	US-08-709-177-76	Sequence 76, Appl	360	5	2.3	144	4	PCT-US93-07213-11	Sequence 11, Appl
288	95	US-09-047-125-30	Sequence 30, Appl	361	5	2.3	144	5	5169835-8	Patent No. 5169835
289	99	US-08-858-767-28	Sequence 28, Appl	362	5	2.3	145	2	US-08-832-535-2	Sequence 2, Appl
290	99	US-08-863-028-28	Sequence 28, Appl	363	5	2.3	146	1	US-08-398-021-3	Sequence 3, Appl
291	99	US-08-710-749-9	Sequence 9, Appl	364	5	2.3	146	1	US-08-678-369-4	Sequence 4, Appl
292	99	US-08-710-749-10	Sequence 10, Appl	365	5	2.3	146	1	US-08-678-369-7	Sequence 7, Appl
293	99	US-08-710-749-11	Sequence 11, Appl	366	5	2.3	146	1	US-08-678-369-8	Sequence 8, Appl
294	99	US-08-710-749-12	Sequence 12, Appl	367	5	2.3	146	1	US-08-788-943A-1	Sequence 1, Appl
295	101	US-08-710-749-1	Sequence 1, Appl	368	5	2.3	146	2	US-08-788-943A-2	Sequence 2, Appl
296	101	US-08-710-749-2	Sequence 2, Appl	369	5	2.3	146	2	US-08-788-943A-3	Sequence 3, Appl
297	101	US-08-710-749-4	Sequence 4, Appl	370	5	2.3	146	2	US-08-788-943A-4	Sequence 4, Appl
298	101	US-08-710-749-7	Sequence 7, Appl	371	5	2.3	146	2	US-08-788-943A-5	Sequence 5, Appl
299	102	US-08-710-749-8	Sequence 8, Appl	372	5	2.3	146	2	US-08-788-943A-6	Sequence 6, Appl
300	104	US-08-315-269-2	Sequence 2, Appl	373	5	2.3	146	2	US-08-788-943A-7	Sequence 7, Appl
301	105	US-08-381-034-1	Sequence 1, Appl	374	5	2.3	146	2	US-08-788-943A-8	Sequence 8, Appl
302	109	US-08-352-324A-2	Sequence 2, Appl	375	5	2.3	146	2	US-08-788-943A-9	Sequence 9, Appl
303	109	US-08-352-324A-9	Sequence 9, Appl	376	5	2.3	146	2	US-08-788-943A-10	Sequence 10, Appl
304	109	US-08-862-607-2	Sequence 2, Appl	377	5	2.3	146	2	US-08-788-943A-11	Sequence 11, Appl
305	109	US-08-862-607-9	Sequence 9, Appl	378	5	2.3	146	2	US-08-788-943A-12	Sequence 12, Appl
306	109	PCT-US95-01780-4	Sequence 4, Appl	379	5	2.3	146	2	US-08-788-943A-13	Sequence 13, Appl
307	109	PCT-US95-16144-2	Sequence 2, Appl	380	5	2.3	146	2	US-08-823-104-3	Sequence 3, Appl
308	109	PCT-US95-16144-9	Sequence 9, Appl	381	5	2.3	146	2	US-08-823-104-4	Sequence 4, Appl
309	110	US-08-362-780-26	Sequence 26, Appl	382	5	2.3	146	2	US-08-823-104-5	Sequence 5, Appl
310	111	US-08-381-050-1	Sequence 1, Appl	383	5	2.3	146	2	US-08-823-104-6	Sequence 6, Appl
311	112	US-08-446-919A-5	Sequence 5, Appl	384	5	2.3	146	2	US-08-823-104-7	Sequence 7, Appl
312	113	US-08-381-057-1	Sequence 1, Appl	385	5	2.3	146	2	US-08-823-104-8	Sequence 8, Appl
313	114	US-08-352-324A-3	Sequence 3, Appl	386	5	2.3	146	2	US-08-823-104-9	Sequence 9, Appl
314	114	US-08-862-607-3	Sequence 3, Appl	387	5	2.3	146	2	US-08-823-104-10	Sequence 10, Appl
315	114	US-08-468-819-4	Sequence 4, Appl	388	5	2.3	146	2	US-08-823-104-11	Sequence 11, Appl
316	114	PCT-US95-16144-3	Sequence 3, Appl	389	5	2.3	146	2	US-08-823-104-12	Sequence 12, Appl
317	115	US-08-111-939-22	Sequence 22, Appl	390	5	2.3	146	2	US-08-823-104-13	Sequence 13, Appl
318	116	US-08-853-659A-42	Sequence 42, Appl	391	5	2.3	146	2	US-08-823-104-14	Sequence 14, Appl
319	117	US-08-557-309B-48	Sequence 48, Appl	392	5	2.3	146	2	US-08-823-104-15	Sequence 15, Appl
320	119	US-08-383-639-1	Sequence 1, Appl	393	5	2.3	146	2	US-08-823-104-16	Sequence 16, Appl

394	5	2.3	146	2	US-09-047-243-4	Sequence 4, Appl	467	5	2.3	189	2	US-09-096-082-3	Sequence 3, Appl
395	5	2.3	146	2	US-09-047-243-7	Sequence 7, Appl	468	5	2.3	192	1	US-08-086-428B-86	Sequence 86, Appl
396	5	2.3	146	2	US-09-047-243-8	Sequence 8, Appl	469	5	2.3	192	1	US-08-086-428B-87	Sequence 87, Appl
397	5	2.3	146	2	US-08-804-668-1	Sequence 1, Appl	470	5	2.3	192	1	US-08-086-428B-88	Sequence 88, Appl
398	5	2.3	146	2	US-09-003-081-6	Sequence 6, Appl	471	5	2.3	192	1	US-08-086-428B-89	Sequence 89, Appl
399	5	2.3	146	2	US-08-648-262-6	Sequence 6, Appl	472	5	2.3	192	2	US-08-468-570-86	Sequence 86, Appl
400	5	2.3	146	3	US-08-648-263-6	Sequence 6, Appl	473	5	2.3	192	2	US-08-468-570-87	Sequence 87, Appl
401	5	2.3	146	4	PCT-US96-00952-1	Sequence 1, Appl	474	5	2.3	192	2	US-08-468-570-88	Sequence 88, Appl
402	5	2.3	146	4	PCT-US96-00952-2	Sequence 2, Appl	475	5	2.3	192	2	US-08-468-570-89	Sequence 89, Appl
403	5	2.3	146	4	PCT-US96-00952-3	Sequence 3, Appl	476	5	2.3	192	2	US-08-290-665A-86	Sequence 86, Appl
404	5	2.3	146	4	PCT-US96-00952-4	Sequence 4, Appl	477	5	2.3	192	2	US-08-290-665A-87	Sequence 87, Appl
405	5	2.3	146	4	PCT-US96-00952-5	Sequence 5, Appl	478	5	2.3	192	2	US-08-290-665A-88	Sequence 88, Appl
406	5	2.3	146	4	PCT-US96-00952-6	Sequence 6, Appl	479	5	2.3	192	2	US-08-290-665A-89	Sequence 89, Appl
407	5	2.3	146	4	PCT-US96-00952-7	Sequence 7, Appl	480	5	2.3	192	4	PCT-US95-10398-86	Sequence 86, Appl
408	5	2.3	146	4	PCT-US96-00952-8	Sequence 8, Appl	481	5	2.3	192	4	PCT-US95-10398-87	Sequence 87, Appl
409	5	2.3	146	4	PCT-US96-00952-9	Sequence 9, Appl	482	5	2.3	192	4	PCT-US95-10398-88	Sequence 88, Appl
410	5	2.3	146	4	PCT-US96-00952-10	Sequence 10, Appl	483	5	2.3	192	4	PCT-US95-10398-89	Sequence 89, Appl
411	5	2.3	146	4	PCT-US96-00952-11	Sequence 11, Appl	484	5	2.3	192	1	US-08-044-621D-29	Sequence 29, Appl
412	5	2.3	148	1	US-08-429-363-3	Sequence 3, Appl	485	5	2.3	197	1	US-08-567-509-1	Sequence 1, Appl
413	5	2.3	148	2	US-08-823-104-17	Sequence 17, Appl	486	5	2.3	197	1	US-08-779-912-9	Sequence 9, Appl
414	5	2.3	148	2	US-08-823-104-18	Sequence 18, Appl	487	5	2.3	197	2	US-08-779-912-9	Sequence 9, Appl
415	5	2.3	152	3	US-08-916-043-2	Sequence 2, Appl	488	5	2.3	197	2	US-08-923-867-2	Sequence 2, Appl
416	5	2.3	152	5	5212286-6	Patent No. 5212286	489	5	2.3	197	2	US-09-047-370-9	Sequence 9, Appl
417	5	2.3	154	2	US-08-756-749C-3	Sequence 3, Appl	490	5	2.3	199	1	US-08-017-522A-4	Sequence 4, Appl
418	5	2.3	155	2	US-08-984-172-3	Sequence 3, Appl	491	5	2.3	199	1	US-08-115-680-2	Sequence 2, Appl
419	5	2.3	157	1	US-08-328-322-15	Sequence 15, Appl	492	5	2.3	199	1	US-07-941-372-2	Sequence 2, Appl
420	5	2.3	158	2	US-08-828-832-3	Sequence 3, Appl	493	5	2.3	199	1	US-08-350-884-67	Sequence 67, Appl
421	5	2.3	158	2	US-08-670-186-6	Sequence 6, Appl	494	5	2.3	199	1	US-07-949-516A-4	Sequence 4, Appl
422	5	2.3	159	2	US-08-162-402B-16	Sequence 16, Appl	495	5	2.3	199	1	US-08-709-173-67	Sequence 67, Appl
423	5	2.3	162	2	US-09-237-507-3	Sequence 3, Appl	496	5	2.3	199	1	US-08-792-019B-8	Sequence 8, Appl
424	5	2.3	164	2	US-08-918-723-3	Sequence 3, Appl	497	5	2.3	199	1	US-08-602-262-2	Sequence 2, Appl
425	5	2.3	165	1	US-08-233-788A-40	Sequence 40, Appl	498	5	2.3	199	2	US-08-814-459-4	Sequence 4, Appl
426	5	2.3	166	2	US-08-918-723-4	Sequence 4, Appl	499	5	2.3	199	2	US-08-709-177-67	Sequence 67, Appl
427	5	2.3	166	2	US-08-347-563A-6	Sequence 6, Appl	500	5	2.3	199	2	PCT-US93-08247-2	Sequence 2, Appl
428	5	2.3	166	2	US-09-237-507-4	Sequence 4, Appl	501	5	2.3	200	2	US-09-031-485-12	Sequence 12, Appl
429	5	2.3	166	3	US-08-292-345B-6	Sequence 6, Appl	502	5	2.3	200	2	US-08-847-429A-12	Sequence 12, Appl
430	5	2.3	167	2	US-08-540-242A-4	Sequence 4, Appl	503	5	2.3	202	1	US-08-350-884-1	Sequence 1, Appl
431	5	2.3	167	2	US-08-347-563A-4	Sequence 4, Appl	504	5	2.3	202	1	US-08-350-884-65	Sequence 65, Appl
432	5	2.3	167	2	US-09-003-081-5	Sequence 5, Appl	505	5	2.3	202	1	US-08-709-173-1	Sequence 1, Appl
433	5	2.3	167	3	US-08-292-345B-4	Sequence 4, Appl	506	5	2.3	202	1	US-08-709-173-65	Sequence 65, Appl
434	5	2.3	167	3	US-08-648-262-5	Sequence 5, Appl	507	5	2.3	202	1	US-08-531-525-24	Sequence 24, Appl
435	5	2.3	167	3	US-08-648-263-5	Sequence 5, Appl	508	5	2.3	202	2	US-08-709-177-1	Sequence 1, Appl
436	5	2.3	167	3	PCT-US96-01471-4	Sequence 4, Appl	509	5	2.3	202	2	US-08-709-177-65	Sequence 65, Appl
437	5	2.3	170	2	US-08-820-170A-22	Sequence 22, Appl	510	5	2.3	202	2	US-08-718-270A-24	Sequence 24, Appl
438	5	2.3	170	2	US-08-742-009-1	Sequence 1, Appl	511	5	2.3	202	2	US-08-598-873-4	Sequence 4, Appl
439	5	2.3	170	2	US-08-918-723-1	Sequence 1, Appl	512	5	2.3	203	2	US-08-446-925-7	Sequence 7, Appl
440	5	2.3	170	2	US-09-237-507-1	Sequence 1, Appl	513	5	2.3	204	2	US-09-146-331-7	Sequence 7, Appl
441	5	2.3	170	3	US-09-172-110-1	Sequence 1, Appl	514	5	2.3	204	2	US-08-696-885-7	Sequence 7, Appl
442	5	2.3	174	1	US-08-328-322-17	Sequence 17, Appl	515	5	2.3	204	2	US-08-576-626A-40	Sequence 40, Appl
443	5	2.3	175	2	US-08-809-267-5	Sequence 5, Appl	516	5	2.3	204	2	US-08-381-881-4	Sequence 4, Appl
444	5	2.3	175	4	PCT-US95-13662A-5	Sequence 5, Appl	517	5	2.3	208	2	US-08-631-607-4	Sequence 4, Appl
445	5	2.3	176	1	US-08-449-922-1	Sequence 1, Appl	518	5	2.3	211	1	US-09-008-303-2	Sequence 2, Appl
446	5	2.3	178	1	US-08-044-621D-32	Sequence 32, Appl	519	5	2.3	213	3	US-08-984-172-1	Sequence 1, Appl
447	5	2.3	178	1	US-07-929-580B-2	Sequence 2, Appl	520	5	2.3	214	2	US-08-997-080-117	Sequence 117, App
448	5	2.3	178	2	US-08-709-912-15	Sequence 15, Appl	521	5	2.3	215	2	US-08-455-543A-43	Sequence 43, Appl
449	5	2.3	178	2	US-09-047-370-15	Sequence 15, Appl	522	5	2.3	216	1	US-08-223-305C-43	Sequence 43, Appl
450	5	2.3	178	2	US-08-791-522-1	Sequence 1, Appl	523	5	2.3	216	2	US-08-149-097D-39	Sequence 39, Appl
451	5	2.3	179	5	5463025-4	Patent No. 5463025	524	5	2.3	216	2	US-07-607-538C-4	Sequence 4, Appl
452	5	2.3	181	1	US-08-049-473-27	Sequence 27, Appl	525	5	2.3	218	1	US-08-126-593A-4	Sequence 4, Appl
453	5	2.3	181	1	US-08-049-473-28	Sequence 28, Appl	526	5	2.3	218	1	US-08-162-402B-4	Sequence 4, Appl
454	5	2.3	181	1	US-08-049-473-29	Sequence 29, Appl	527	5	2.3	218	1	US-08-454-039A-4	Sequence 4, Appl
455	5	2.3	181	1	US-08-312-648-27	Sequence 27, Appl	528	5	2.3	218	2	US-08-455-543A-44	Sequence 44, Appl
456	5	2.3	181	1	US-08-312-648-28	Sequence 28, Appl	529	5	2.3	219	2	US-08-223-305C-44	Sequence 44, Appl
457	5	2.3	181	1	US-08-312-648-29	Sequence 29, Appl	530	5	2.3	223	2	US-08-149-097D-40	Sequence 40, Appl
458	5	2.3	181	1	US-08-418-444A-4	Sequence 4, Appl	531	5	2.3	224	4	US-08-896-410-4	Sequence 4, Appl
459	5	2.3	181	1	US-08-418-444A-5	Sequence 5, Appl	532	5	2.3	224	4	PCT-US91-09055-4	Sequence 5, Appl
460	5	2.3	181	1	PCT-US94-04190-27	Sequence 27, Appl	533	5	2.3	224	1	US-08-414-625-4	Sequence 4, Appl
461	5	2.3	181	4	PCT-US94-04190-28	Sequence 28, Appl	534	5	2.3	227	1	US-08-278-091-7	Sequence 7, Appl
462	5	2.3	181	4	PCT-US94-04190-29	Sequence 29, Appl	535	5	2.3	228	1	US-08-483-859-7	Sequence 7, Appl
463	5	2.3	182	2	US-08-558-823-5	Sequence 5, Appl	536	5	2.3	228	1	US-08-472-173-7	Sequence 7, Appl
464	5	2.3	186	1	US-08-117-083-22	Sequence 22, Appl	537	5	2.3	228	1		
465	5	2.3	189	1	US-08-884-682-3	Sequence 3, Appl	538	5	2.3				
466	5	2.3					539	5	2.3				

540	5	2.3	228	2	US-08-487-167-7	Sequence 7, Appl	613	5	2.3	271	3	US-08-536-891A-1	Sequence 1, Appl
541	5	2.3	228	2	US-08-482-816-7	Sequence 7, Appl	614	5	2.3	272	1	US-08-350-884-84	Sequence 84, Appl
542	5	2.3	228	2	US-08-296-149-7	Sequence 7, Appl	615	5	2.3	272	1	US-08-709-173-84	Sequence 84, Appl
543	5	2.3	228	2	US-08-801-499-7	Sequence 7, Appl	616	5	2.3	272	2	US-08-852-743-6	Sequence 6, Appl
544	5	2.3	228	2	US-08-615-271-7	Sequence 7, Appl	617	5	2.3	272	2	US-08-160-524A-7	Sequence 7, Appl
545	5	2.3	228	3	US-09-074-660-7	Sequence 7, Appl	618	5	2.3	272	2	US-08-709-177-84	Sequence 84, Appl
546	5	2.3	228	3	US-09-074-659-7	Sequence 7, Appl	619	5	2.3	273	1	US-08-215-928A-2	Sequence 2, Appl
547	5	2.3	229	1	US-08-028-463-2	Sequence 2, Appl	620	5	2.3	273	1	US-08-446-920-2	Sequence 2, Appl
548	5	2.3	229	1	US-08-383-750-6	Sequence 6, Appl	621	5	2.3	275	2	US-08-701-131A-38	Sequence 38, Appl
549	5	2.3	229	1	US-08-461-836-2	Sequence 2, Appl	622	5	2.3	276	2	US-08-070-301-11	Sequence 11, Appl
550	5	2.3	229	2	US-08-684-687-2	Sequence 2, Appl	623	5	2.3	277	2	US-08-469-537A-57	Sequence 57, Appl
551	5	2.3	229	2	US-08-121-436A-4	Sequence 4, Appl	624	5	2.3	278	1	US-07-941-414-1	Sequence 1, Appl
552	5	2.3	229	2	US-08-801-740-3	Sequence 3, Appl	625	5	2.3	278	1	US-08-249-377A-1	Sequence 1, Appl
553	5	2.3	229	3	US-08-352-678-6	Sequence 6, Appl	626	5	2.3	278	1	US-08-585-595-1	Sequence 1, Appl
554	5	2.3	230	1	US-08-052-205-11	Sequence 11, Appl	627	5	2.3	278	2	US-08-416-603-6	Sequence 6, Appl
555	5	2.3	230	1	US-08-595-974-11	Sequence 11, Appl	628	5	2.3	278	2	US-08-749-337-1	Sequence 1, Appl
556	5	2.3	230	2	US-08-637-759B-457	Sequence 457, App	629	5	2.3	278	2	US-08-865-203-6	Sequence 6, Appl
557	5	2.3	230	3	US-08-871-355A-457	Sequence 457, App	630	5	2.3	278	2	US-07-849-420-6	Sequence 6, Appl
558	5	2.3	230	4	PCT-US93-09636-6	Sequence 6, Appl	631	5	2.3	278	3	US-08-522-813-4	Sequence 4, Appl
559	5	2.3	231	2	US-08-902-655A-6	Sequence 6, Appl	632	5	2.3	278	4	PCT-US94-03744-2	Sequence 2, Appl
560	5	2.3	232	2	US-08-446-345-38	Sequence 38, Appl	633	5	2.3	279	2	US-09-135-658-4	Sequence 4, Appl
561	5	2.3	233	1	US-08-800-751-2	Sequence 2, Appl	634	5	2.3	279	4	PCT-US91-00899-8	Sequence 8, Appl
562	5	2.3	233	2	US-08-738-413B-9	Sequence 9, Appl	635	5	2.3	279	5	5472855-6	Patent No. 5472855
563	5	2.3	233	2	US-08-990-818-2	Sequence 2, Appl	636	5	2.3	280	1	US-08-312-387B-6	Sequence 6, Appl
564	5	2.3	234	1	US-08-299-567-5	Sequence 5, Appl	637	5	2.3	280	1	US-08-683-426-6	Sequence 6, Appl
565	5	2.3	236	1	US-08-462-949-3	Sequence 3, Appl	638	5	2.3	280	1	US-08-683-458-6	Sequence 6, Appl
566	5	2.3	236	1	US-08-023-764B-3	Sequence 3, Appl	639	5	2.3	280	2	US-08-878-360-6	Sequence 6, Appl
567	5	2.3	237	2	US-08-103-170-11	Sequence 11, Appl	640	5	2.3	284	2	US-08-767-096-3	Sequence 3, Appl
568	5	2.3	238	1	US-08-240-124-2	Sequence 2, Appl	641	5	2.3	285	1	US-08-631-200-15	Sequence 15, Appl
569	5	2.3	238	1	US-08-453-943-2	Sequence 2, Appl	642	5	2.3	285	2	US-08-829-553-15	Sequence 15, Appl
570	5	2.3	238	2	US-09-057-121-2	Sequence 2, Appl	643	5	2.3	285	2	US-08-922-267A-15	Sequence 15, Appl
571	5	2.3	248	5	5169835-15	Patent No. 5169835	644	5	2.3	285	2	US-08-936-707A-15	Sequence 15, Appl
572	5	2.3	248	5	5182262-8	Patent No. 5182262	645	5	2.3	285	2	US-08-936-706A-15	Sequence 15, Appl
573	5	2.3	249	3	US-08-486-099-102	Sequence 102, App	646	5	2.3	285	3	US-09-248-203-15	Sequence 15, Appl
574	5	2.3	249	3	US-08-484-223B-102	Sequence 102, App	647	5	2.3	287	1	US-08-622-353-7	Sequence 7, Appl
575	5	2.3	250	3	US-08-360-107A-112	Sequence 112, App	648	5	2.3	287	2	US-08-622-352A-9	Sequence 9, Appl
576	5	2.3	251	2	US-08-469-537A-80	Sequence 80, Appl	649	5	2.3	290	2	US-08-576-626A-34	Sequence 34, Appl
577	5	2.3	251	2	US-08-469-537A-81	Sequence 81, Appl	650	5	2.3	292	1	US-08-286-888B-6	Sequence 6, Appl
578	5	2.3	251	4	PCT-US96-01314-59	Sequence 59, Appl	651	5	2.3	293	1	US-08-446-945-5	Sequence 5, Appl
579	5	2.3	252	1	US-08-052-205-9	Sequence 9, Appl	652	5	2.3	293	2	US-08-896-885-5	Sequence 5, Appl
580	5	2.3	252	1	US-08-595-974-9	Sequence 9, Appl	653	5	2.3	293	4	PCT-US96-10521-31	Sequence 31, Appl
581	5	2.3	252	1	US-08-460-512-4	Sequence 4, Appl	654	5	2.3	294	3	US-09-057-969-7	Sequence 7, Appl
582	5	2.3	252	2	US-08-853-659A-43	Sequence 43, Appl	655	5	2.3	294	3	US-07-745-382-14	Sequence 14, Appl
583	5	2.3	255	2	US-08-685-992-16	Sequence 16, Appl	656	5	2.3	296	1	US-07-921-848-14	Sequence 14, Appl
584	5	2.3	255	2	US-09-144-923-16	Sequence 16, Appl	657	5	2.3	296	1	US-08-115-680-4	Sequence 4, Appl
585	5	2.3	255	2	PCT-US95-04570-106	Sequence 106, App	658	5	2.3	296	1	US-07-941-372-4	Sequence 4, Appl
586	5	2.3	255	4	PCT-US95-04589-106	Sequence 106, App	659	5	2.3	296	1	US-08-165-301A-14	Sequence 14, Appl
587	5	2.3	255	4	PCT-US95-04589-106	Sequence 106, App	660	5	2.3	296	1	US-08-507-431-6	Sequence 6, Appl
588	5	2.3	257	1	US-08-467-265-16	Sequence 16, Appl	661	5	2.3	296	2	US-08-836-854-4	Sequence 4, Appl
589	5	2.3	258	1	US-08-744-026-3	Sequence 3, Appl	662	5	2.3	296	2	PCT-US93-08247-4	Sequence 4, Appl
590	5	2.3	258	2	US-09-102-732-3	Sequence 3, Appl	663	5	2.3	296	4	PCT-US94-14179-14	Sequence 14, Appl
591	5	2.3	259	2	US-08-402-804-8	Sequence 8, Appl	664	5	2.3	297	1	US-07-866-560-6	Sequence 6, Appl
592	5	2.3	260	2	US-08-172-977-1	Sequence 1, Appl	665	5	2.3	297	1	US-08-077-673-6	Sequence 6, Appl
593	5	2.3	262	1	US-08-738-944-50	Sequence 50, Appl	666	5	2.3	297	1	US-08-478-992-6	Sequence 6, Appl
594	5	2.3	262	1	US-08-744-026-4	Sequence 4, Appl	667	5	2.3	299	1	US-08-350-884-66	Sequence 66, Appl
595	5	2.3	262	2	US-08-790-137-1	Sequence 1, Appl	668	5	2.3	299	1	US-08-350-884-68	Sequence 68, Appl
596	5	2.3	262	2	US-08-790-137-3	Sequence 3, Appl	669	5	2.3	299	1	US-08-709-173-66	Sequence 66, Appl
597	5	2.3	262	2	US-08-681-151-4	Sequence 4, Appl	670	5	2.3	299	1	US-08-709-173-68	Sequence 68, Appl
598	5	2.3	262	2	US-09-102-732-4	Sequence 4, Appl	671	5	2.3	299	2	US-08-709-177-66	Sequence 66, Appl
599	5	2.3	262	2	US-07-857-224B-62	Sequence 62, Appl	672	5	2.3	299	2	US-08-709-177-68	Sequence 68, Appl
600	5	2.3	262	2	US-08-824-874-4	Sequence 4, Appl	673	5	2.3	303	1	US-07-917-111-2	Sequence 2, Appl
601	5	2.3	262	3	US-08-807-151-4	Sequence 4, Appl	674	5	2.3	303	1	US-08-479-638-2	Sequence 2, Appl
602	5	2.3	263	2	US-08-790-137-4	Sequence 4, Appl	675	5	2.3	303	1	US-08-294-871A-8	Sequence 8, Appl
603	5	2.3	263	2	US-08-824-874-5	Sequence 5, Appl	676	5	2.3	303	1	US-08-294-871A-10	Sequence 10, Appl
604	5	2.3	263	3	US-09-035-706-5	Sequence 5, Appl	677	5	2.3	303	1	US-08-294-871A-12	Sequence 12, Appl
605	5	2.3	263	3	US-08-955-841-5	Sequence 5, Appl	678	5	2.3	303	2	US-08-294-871A-14	Sequence 14, Appl
606	5	2.3	263	3	US-08-807-151-5	Sequence 5, Appl	679	5	2.3	303	2	US-08-294-871A-16	Sequence 16, Appl
607	5	2.3	266	5	5175384-11	Patent No. 5175384	680	5	2.3	303	2	US-08-294-871A-6	Sequence 6, Appl
608	5	2.3	267	2	US-07-857-224B-37	Sequence 37, Appl	681	5	2.3	303	2	US-08-294-871A-8	Sequence 8, Appl
609	5	2.3	267	2	US-08-793-490-11	Sequence 11, Appl	682	5	2.3	303	2	US-08-294-871A-10	Sequence 10, Appl
610	5	2.3	268	3	US-08-836-134-22	Sequence 22, Appl	683	5	2.3	303	2	US-08-294-871A-12	Sequence 12, Appl
611	5	2.3	269	1	US-08-241-766-6	Sequence 6, Appl	684	5	2.3	303	2	US-08-294-871A-14	Sequence 14, Appl
612	5	2.3	269	1	US-08-241-766-7	Sequence 7, Appl	685	5	2.3	303	2	US-08-294-871A-16	Sequence 16, Appl

686	2.3	303	2	US-08-294-871A-18	Sequence 18, Appl	759	2.3	332	2	US-08-637-763B-8	Sequence 8, Appl
687	2.3	303	2	US-08-294-871A-20	Sequence 20, Appl	760	2.3	333	2	US-08-853-659A-48	Sequence 48, Appl
688	2.3	303	2	US-08-294-871A-22	Sequence 22, Appl	761	2.3	334	2	US-08-559-524A-2	Sequence 2, Appl
689	2.3	303	2	US-08-294-871A-24	Sequence 24, Appl	762	2.3	334	3	US-08-472-595-9	Sequence 9, Appl
690	2.3	303	2	US-08-294-871A-26	Sequence 26, Appl	763	2.3	334	3	US-08-207-575A-9	Sequence 9, Appl
691	2.3	303	2	US-08-294-871A-28	Sequence 28, Appl	764	2.3	336	2	US-08-997-080-156	Sequence 156, App
692	2.3	303	2	US-08-294-871A-30	Sequence 30, Appl	765	2.3	336	2	US-08-997-362-156	Sequence 156, App
693	2.3	303	2	US-08-294-871A-32	Sequence 32, Appl	766	2.3	337	2	US-08-118-270-33	Sequence 33, Appl
694	2.3	303	2	US-08-294-871A-34	Sequence 34, Appl	767	2.3	337	2	US-08-467-559B-2	Sequence 2, Appl
695	2.3	303	2	US-08-294-871A-36	Sequence 36, Appl	768	2.3	337	4	PCT-US93-08528-33	Sequence 40, Appl
696	2.3	303	2	US-08-294-871A-38	Sequence 38, Appl	769	2.3	338	2	US-08-602-359A-40	Sequence 33, Appl
697	2.3	303	2	US-08-294-871A-40	Sequence 40, Appl	770	2.3	339	1	US-08-396-957A-4	Sequence 4, Appl
698	2.3	303	2	US-08-294-871A-42	Sequence 42, Appl	771	2.3	344	1	US-08-400-422-4	Sequence 4, Appl
699	2.3	303	2	US-08-294-871A-44	Sequence 44, Appl	772	2.3	344	1	US-08-843-993-3	Sequence 3, Appl
700	2.3	303	2	US-08-294-871A-46	Sequence 46, Appl	773	2.3	344	1	US-08-657-192-3	Sequence 1, Appl
701	2.3	303	2	US-08-294-871A-48	Sequence 48, Appl	774	2.3	344	1	US-08-689-974-1	Sequence 1, Appl
702	2.3	303	2	US-08-294-871A-50	Sequence 50, Appl	775	2.3	344	3	US-08-625-377-4	Sequence 4, Appl
703	2.3	303	2	US-08-294-871A-52	Sequence 52, Appl	776	2.3	344	3	US-09-059-520A-3	Sequence 3, Appl
704	2.3	303	2	US-08-294-871A-54	Sequence 54, Appl	777	2.3	344	3	US-09-334-275-3	Sequence 3, Appl
705	2.3	303	2	US-08-294-871A-56	Sequence 56, Appl	778	2.3	344	3	US-08-523-373-5	Sequence 5, Appl
706	2.3	303	2	US-08-294-871A-58	Sequence 58, Appl	779	2.3	344	5	5210183-2	Patent No. 5210183
707	2.3	303	2	US-08-294-871A-60	Sequence 60, Appl	780	2.3	345	1	US-08-171-382-6	Sequence 6, Appl
708	2.3	303	2	US-08-294-871A-62	Sequence 62, Appl	781	2.3	345	1	US-08-118-270-43	Sequence 43, Appl
709	2.3	303	2	US-08-294-871A-64	Sequence 64, Appl	782	2.3	345	1	US-08-309-420-6	Sequence 6, Appl
710	2.3	303	2	US-08-294-871A-66	Sequence 66, Appl	783	2.3	345	1	US-08-309-419-6	Sequence 6, Appl
711	2.3	303	2	US-08-294-871A-68	Sequence 68, Appl	784	2.3	345	1	US-08-843-993-1	Sequence 1, Appl
712	2.3	303	2	US-08-203-532B-2	Sequence 2, Appl	785	2.3	345	4	PCT-US93-08528-43	Sequence 43, Appl
713	2.3	303	4	PCT-US95-01882A-2	Sequence 2, Appl	786	2.3	345	4	PCT-US95-11856-6	Sequence 6, Appl
714	2.3	304	1	US-08-900-711-5	Sequence 5, Appl	787	2.3	345	3	US-09-059-520A-1	Sequence 3, Appl
715	2.3	304	2	US-08-815-356-2	Sequence 5, Appl	788	2.3	345	3	US-09-334-275-1	Sequence 1, Appl
716	2.3	304	2	US-08-415-343B-5	Sequence 5, Appl	789	2.3	345	4	PCT-US93-08528-43	Sequence 43, Appl
717	2.3	304	2	US-08-762-433-1	Sequence 5, Appl	790	2.3	345	4	PCT-US95-11856-6	Sequence 6, Appl
718	2.3	304	2	US-09-001-219-1	Sequence 1, Appl	791	2.3	346	1	US-08-774-169-3	Sequence 1, Appl
719	2.3	307	2	US-08-807-861A-56	Sequence 1, Appl	792	2.3	347	1	US-08-052-205-7	Sequence 7, Appl
720	2.3	308	1	US-08-457-245-8	Sequence 56, Appl	793	2.3	347	1	US-08-446-920-10	Sequence 10, Appl
721	2.3	309	2	US-08-405-175A-7	Sequence 8, Appl	794	2.3	347	2	US-08-595-974-7	Sequence 7, Appl
722	2.3	309	2	US-08-405-175A-8	Sequence 8, Appl	795	2.3	347	2	US-08-773-870-3	Sequence 3, Appl
723	2.3	310	1	US-08-078-683A-3	Sequence 8, Appl	796	2.3	347	3	US-09-059-369-2	Sequence 2, Appl
724	2.3	310	1	US-08-433-783-42	Sequence 3, Appl	797	2.3	348	3	US-08-875-540-13	Sequence 13, Appl
725	2.3	310	2	US-08-484-956-88	Sequence 42, Appl	798	2.3	349	1	US-08-872-784-4	Sequence 4, Appl
726	2.3	310	2	US-08-484-956-88	Sequence 42, Appl	799	2.3	349	2	US-08-558-823-4	Sequence 4, Appl
727	2.3	310	2	US-08-757-653-88	Sequence 42, Appl	800	2.3	349	2	US-09-100-851-4	Sequence 4, Appl
728	2.3	310	2	US-08-488-199-4	Sequence 42, Appl	801	2.3	351	1	US-08-159-969-2	Sequence 2, Appl
729	2.3	310	4	PCT-US95-07537A-42	Sequence 42, Appl	802	2.3	351	1	US-08-552-142A-4	Sequence 4, Appl
730	2.3	310	4	PCT-US95-07537-42	Sequence 42, Appl	803	2.3	352	1	US-08-910-973-4	Sequence 4, Appl
731	2.3	312	1	US-08-414-625-2	Sequence 2, Appl	804	2.3	352	2	US-08-751-474-2	Sequence 2, Appl
732	2.3	314	5	5434340-7	Patent No. 5434340	805	2.3	352	2	US-08-726-306A-17	Sequence 17, Appl
733	2.3	315	2	US-08-484-956-91	Sequence 91, Appl	806	2.3	352	4	PCT-US95-05741-4	Sequence 4, Appl
734	2.3	315	2	US-08-757-653-91	Sequence 91, Appl	807	2.3	354	2	US-08-700-013B-9	Sequence 9, Appl
735	2.3	317	2	US-08-864-799-4	Sequence 4, Appl	808	2.3	354	2	US-08-394-189B-20	Sequence 20, Appl
736	2.3	317	2	US-08-864-799-5	Sequence 5, Appl	809	2.3	355	1	US-08-461-244-2	Sequence 2, Appl
737	2.3	317	2	US-08-377-847-3	Sequence 3, Appl	810	2.3	355	2	US-08-666-367B-6	Sequence 6, Appl
738	2.3	317	2	US-09-195-021-3	Sequence 3, Appl	811	2.3	356	1	US-07-959-941-2	Sequence 2, Appl
739	2.3	320	2	US-08-757-653-163	Sequence 163, App	812	2.3	356	1	US-08-259-924-2	Sequence 2, Appl
740	2.3	320	2	US-08-923-516-61	Sequence 61, Appl	813	2.3	358	2	US-08-748-485-6	Sequence 6, Appl
741	2.3	321	1	US-08-447-185-1	Sequence 1, Appl	814	2.3	359	1	US-07-996-772A-10	Sequence 10, Appl
742	2.3	322	2	US-08-484-956-89	Sequence 89, Appl	815	2.3	359	2	US-08-467-568-13	Sequence 13, Appl
743	2.3	322	2	US-08-757-653-89	Sequence 89, Appl	816	2.3	359	2	US-08-748-485-3	Sequence 3, Appl
744	2.3	322	2	US-08-576-626A-33	Sequence 33, Appl	817	2.3	359	2	US-08-748-485-4	Sequence 4, Appl
745	2.3	323	2	US-08-540-804-18	Sequence 18, Appl	818	2.3	359	2	US-08-748-485-5	Sequence 5, Appl
746	2.3	323	2	US-08-218-265-18	Sequence 18, Appl	819	2.3	359	2	US-08-103-170-2	Sequence 2, Appl
747	2.3	323	2	US-08-521-872-18	Sequence 18, Appl	820	2.3	359	2	US-08-103-170-4	Sequence 4, Appl
748	2.3	327	2	US-08-896-410-38	Sequence 38, Appl	821	2.3	359	2	US-08-103-170-6	Sequence 6, Appl
749	2.3	327	2	US-08-926-724-1	Sequence 1, Appl	822	2.3	359	2	US-08-103-170-7	Sequence 7, Appl
750	2.3	327	2	US-08-926-724-3	Sequence 3, Appl	823	2.3	359	2	US-09-030-583-13	Sequence 13, Appl
751	2.3	328	1	US-08-148-215A-2	Sequence 2, Appl	824	2.3	359	3	US-08-875-540-15	Sequence 15, Appl
752	2.3	328	2	US-08-928-242-1	Sequence 2, Appl	825	2.3	360	1	US-08-118-270-25	Sequence 25, Appl
753	2.3	328	2	US-08-910-927B-1	Sequence 1, Appl	826	2.3	360	4	PCT-US93-08528-25	Sequence 25, Appl
754	2.3	329	2	US-08-606-143-1	Sequence 1, Appl	827	2.3	361	1	US-08-415-751-3	Sequence 3, Appl
755	2.3	329	2	US-08-606-143-2	Sequence 2, Appl	828	2.3	362	1	US-08-466-103A-16	Sequence 16, Appl
756	2.3	329	2	US-08-606-143-3	Sequence 3, Appl	829	2.3	363	2	US-08-530-950-6	Sequence 6, Appl
757	2.3	331	2	US-08-997-080-182	Sequence 182, App	830	2.3	364	1	US-08-318-831-6	Sequence 6, Appl
758	2.3	331	2	US-08-997-362-182	Sequence 182, App	831	2.3	364	1	US-08-650-275-4	Sequence 4, Appl

832	5	2.3	364	3	US-09-181-318-4	Sequence 4, Appl1	905	5	2.3	406	2	US-09-203-237-5	Sequence 5, Appl1
833	5	2.3	365	2	US-08-855-518-5	Sequence 5, Appl1	906	5	2.3	406	5	5212236-6	Patent No. 5212236
834	5	2.3	365	2	US-08-169-948B-12	Sequence 12, Appl	907	5	2.3	411	1	US-08-381-433A-4	Sequence 4, Appl1
835	5	2.3	365	2	US-08-448-873-12	Sequence 12, Appl	908	5	2.3	411	1	US-08-381-433A-8	Sequence 8, Appl1
836	5	2.3	365	2	US-08-979-424-3	Sequence 3, Appl1	909	5	2.3	411	1	US-08-399-561-2	Sequence 2, Appl1
837	5	2.3	366	1	US-08-774-169-1	Sequence 1, Appl1	910	5	2.3	414	2	US-08-750-524-1	Sequence 1, Appl1
838	5	2.3	366	2	US-08-855-518-4	Sequence 4, Appl1	911	5	2.3	415	1	US-08-110-286A-2	Sequence 2, Appl1
839	5	2.3	367	2	US-08-896-320-1	Sequence 1, Appl1	912	5	2.3	415	1	US-08-110-286A-6	Sequence 6, Appl1
840	5	2.3	367	2	US-08-896-320-3	Sequence 3, Appl1	913	5	2.3	416	4	PCT-US95-15463-17	Sequence 17, Appl
841	5	2.3	367	3	US-09-032-372-4	Sequence 4, Appl1	914	5	2.3	416	4	PCT-US95-15923-17	Sequence 17, Appl
842	5	2.3	369	1	US-08-052-205-4	Sequence 4, Appl1	915	5	2.3	419	2	US-08-997-080-125	Sequence 125, App
843	5	2.3	369	1	US-08-595-974-4	Sequence 4, Appl1	916	5	2.3	419	5	US-08-997-362-125	Sequence 125, App
844	5	2.3	371	2	US-08-378-617A-12	Sequence 12, Appl	917	5	2.3	420	1	US-08-997-362-125	Patent No. 5169835
845	5	2.3	372	1	US-07-813-584A-3	Sequence 3, Appl1	918	5	2.3	420	1	US-08-487-823B-4	Sequence 4, Appl1
846	5	2.3	372	1	US-08-330-515-3	Sequence 3, Appl1	919	5	2.3	420	2	US-08-997-040-4	Sequence 4, Appl1
847	5	2.3	372	2	US-08-846-762-79	Sequence 79, Appl	920	5	2.3	420	2	US-09-203-237-4	Sequence 4, Appl1
848	5	2.3	373	1	US-08-118-270-24	Sequence 24, Appl	921	5	2.3	423	2	US-08-290-731C-10	Sequence 10, Appl
849	5	2.3	373	4	PCT-US93-08528-24	Sequence 24, Appl	922	5	2.3	424	5	US-08-997-080-125	Patent No. 5169835
850	5	2.3	374	1	US-08-095-726-14	Sequence 14, Appl	923	5	2.3	425	2	US-08-732-028-2	Sequence 2, Appl1
851	5	2.3	374	1	US-08-096-623A-14	Sequence 14, Appl	924	5	2.3	425	2	US-08-732-028-2	Sequence 2, Appl1
852	5	2.3	376	2	US-08-737-825-6	Sequence 6, Appl1	925	5	2.3	428	1	US-08-570-157-5	Sequence 5, Appl1
853	5	2.3	376	5	5180810-1	Sequence 6, Appl1	926	5	2.3	428	5	5179023-4	Patent No. 5179023
854	5	2.3	378	1	US-07-772-087-3	Sequence 3, Appl1	927	5	2.3	429	1	US-07-602-824A-2	Sequence 2, Appl1
855	5	2.3	378	3	US-09-057-969-6	Sequence 6, Appl1	928	5	2.3	429	1	US-07-602-824A-2	Sequence 2, Appl1
856	5	2.3	380	1	US-08-420-235B-5	Sequence 5, Appl1	929	5	2.3	429	1	US-07-983-451-2	Sequence 3, Appl1
857	5	2.3	380	1	PCT-US95-10194-5	Sequence 5, Appl1	930	5	2.3	429	1	US-08-261-578-3	Sequence 3, Appl1
858	5	2.3	381	2	US-08-858-052-1	Sequence 1, Appl1	931	5	2.3	429	1	US-08-261-577-9	Sequence 9, Appl1
859	5	2.3	381	2	US-08-858-052-3	Sequence 3, Appl1	932	5	2.3	430	1	US-07-937-609-23	Sequence 23, Appl
860	5	2.3	383	1	US-08-464-523B-31	Sequence 31, Appl	933	5	2.3	430	2	US-08-919-624-3	Sequence 3, Appl1
861	5	2.3	385	1	US-08-450-257-58	Sequence 58, Appl	934	5	2.3	431	1	US-08-190-802A-37	Sequence 37, Appl
862	5	2.3	385	1	US-08-450-246-58	Sequence 58, Appl	935	5	2.3	431	1	US-08-381-433A-2	Sequence 2, Appl1
863	5	2.3	385	1	US-08-450-098-58	Sequence 58, Appl	936	5	2.3	431	2	US-08-665-647-5	Sequence 5, Appl1
864	5	2.3	385	1	US-08-451-233-58	Sequence 58, Appl	937	5	2.3	433	1	US-08-810-116-9	Sequence 9, Appl1
865	5	2.3	385	1	US-08-451-233-58	Sequence 58, Appl	938	5	2.3	433	3	US-07-930-548A-9	Sequence 9, Appl1
866	5	2.3	388	1	US-08-261-110A-4	Sequence 4, Appl1	939	5	2.3	433	3	US-08-691-563C-87	Sequence 87, Appl
867	5	2.3	389	2	US-08-820-170A-13	Sequence 13, Appl	940	5	2.3	435	1	US-08-374-155A-5	Sequence 5, Appl1
868	5	2.3	390	1	US-07-669-171-2	Sequence 2, Appl1	941	5	2.3	435	2	US-08-785-396-5	Sequence 5, Appl1
869	5	2.3	390	1	US-07-937-609-18	Sequence 18, Appl	942	5	2.3	436	1	US-08-146-010A-2	Sequence 2, Appl1
870	5	2.3	390	1	US-08-132-405-1	Sequence 1, Appl1	943	5	2.3	436	1	US-08-674-168-2	Sequence 2, Appl1
871	5	2.3	390	1	US-08-395-939A-1	Sequence 1, Appl1	944	5	2.3	436	1	US-08-674-168-6	Sequence 6, Appl1
872	5	2.3	390	4	PCT-US91-01861-1	Sequence 1, Appl1	945	5	2.3	438	1	US-08-261-110A-2	Sequence 2, Appl1
873	5	2.3	390	4	PCT-US92-02091-6	Sequence 6, Appl1	946	5	2.3	438	4	PCT-US92-09325-4	Sequence 4, Appl1
874	5	2.3	390	5	5168051-2	Sequence 5, Appl1	947	5	2.3	439	3	US-08-993-359-24	Sequence 24, Appl
875	5	2.3	391	2	US-08-444-951A-10	Sequence 10, Appl	948	5	2.3	440	1	US-08-307-499-15	Sequence 15, Appl
876	5	2.3	391	2	US-08-389-011-23	Sequence 23, Appl	949	5	2.3	442	1	US-08-208-108-2	Sequence 2, Appl1
877	5	2.3	391	3	US-08-403-917A-23	Sequence 23, Appl	950	5	2.3	442	3	US-08-993-359-26	Sequence 26, Appl
878	5	2.3	392	1	US-08-423-441-2	Sequence 2, Appl1	951	5	2.3	442	3	US-08-993-359-28	Sequence 28, Appl
879	5	2.3	392	3	US-08-523-373-6	Sequence 6, Appl1	952	5	2.3	443	2	US-08-570-157-6	Sequence 6, Appl1
880	5	2.3	393	2	US-08-530-950-8	Sequence 8, Appl1	953	5	2.3	443	2	US-08-795-475-6	Sequence 6, Appl1
881	5	2.3	393	2	US-08-530-379-4	Sequence 4, Appl1	954	5	2.3	443	3	US-08-993-359-30	Sequence 30, Appl
882	5	2.3	394	1	US-07-914-281-4	Sequence 4, Appl1	955	5	2.3	444	1	US-07-937-609-14	Sequence 14, Appl
883	5	2.3	394	1	US-08-393-246-4	Sequence 4, Appl1	956	5	2.3	445	1	US-08-700-359-4	Sequence 4, Appl1
884	5	2.3	394	2	US-08-525-058A-4	Sequence 4, Appl1	957	5	2.3	445	4	PCT-US92-09325-2	Sequence 2, Appl1
885	5	2.3	394	2	US-08-696-731-4	Sequence 4, Appl1	958	5	2.3	446	1	US-07-626-618A-21	Sequence 21, Appl
886	5	2.3	394	2	PCT-US91-00899-9	Sequence 9, Appl1	959	5	2.3	446	1	US-07-626-618A-22	Sequence 22, Appl
887	5	2.3	394	4	PCT-US94-03705-5	Sequence 5, Appl1	960	5	2.3	446	1	US-08-333-977-21	Sequence 21, Appl
888	5	2.3	395	1	US-08-357-264-1	Sequence 1, Appl1	961	5	2.3	446	1	US-08-333-977-22	Sequence 22, Appl
889	5	2.3	395	1	US-08-672-514-1	Sequence 1, Appl1	962	5	2.3	446	2	US-07-969-267B-4	Sequence 4, Appl1
890	5	2.3	395	2	US-08-990-379-5	Sequence 5, Appl1	963	5	2.3	446	2	US-08-833-610-2	Sequence 2, Appl1
891	5	2.3	395	2	US-08-990-379-5	Sequence 5, Appl1	964	5	2.3	447	1	US-08-476-008-67	Sequence 67, Appl
892	5	2.3	399	1	US-08-530-950-10	Sequence 10, Appl	965	5	2.3	447	1	US-08-306-063-67	Sequence 67, Appl
893	5	2.3	399	2	US-08-874-186-92	Sequence 92, Appl	966	5	2.3	447	1	US-08-833-485-67	Sequence 67, Appl
894	5	2.3	401	3	US-08-974-022-2	Sequence 2, Appl1	967	5	2.3	447	3	US-08-508-761B-6	Sequence 6, Appl1
895	5	2.3	402	2	US-08-068-729-4	Sequence 4, Appl1	968	5	2.3	450	1	US-07-937-609-24	Sequence 24, Appl
896	5	2.3	404	2	US-08-947-726A-2	Sequence 2, Appl1	969	5	2.3	451	1	US-08-435-454-4	Sequence 4, Appl1
897	5	2.3	405	1	US-07-829-954-2	Sequence 2, Appl1	970	5	2.3	451	2	US-08-652-972A-4	Sequence 4, Appl1
898	5	2.3	405	1	US-07-994-423-2	Sequence 2, Appl1	971	5	2.3	451	2	US-08-919-145-6	Sequence 6, Appl1
899	5	2.3	405	1	US-08-421-891-2	Sequence 2, Appl1	972	5	2.3	451	4	PCT-US96-06231A-4	Sequence 4, Appl1
900	5	2.3	405	1	US-08-406-070-2	Sequence 2, Appl1	973	5	2.3	452	1	US-08-317-880-2	Sequence 2, Appl1
901	5	2.3	405	2	US-08-282-197C-61	Sequence 61, Appl	974	5	2.3	452	2	US-08-782-396-2	Sequence 2, Appl1
902	5	2.3	405	2	US-08-700-013B-6	Sequence 6, Appl1	975	5	2.3	453	3	US-08-993-359-22	Sequence 22, Appl
903	5	2.3	406	1	US-08-487-823B-5	Sequence 5, Appl1	976	5	2.3	454	1	US-08-171-382-4	Sequence 4, Appl1
904	5	2.3	406	2	US-08-997-040-5	Sequence 5, Appl1	977	5	2.3	454	1	US-08-509-420-4	Sequence 4, Appl1

Sequence 4, Appl1
Sequence 9, Appl1
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 3, Appl1
Sequence 5, Appl1
Sequence 3, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 5, Appl1
Sequence 9, Appl1
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 1, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 9, Appl1
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 2, Appl1
Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-225-488-7
; Sequence 7, Application US/08225488
; Patent No. 5846802
; GENERAL INFORMATION:
; APPLICANT: Buxton, Frank
; APPLICANT: Hinnen, Albert
; APPLICANT: Visser, Jacob
; TITLE OF INVENTION: No. 5846802el Fungal Protease
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/225,488
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 08/047,214
; FILING DATE: 13-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 4-19055/A/CIP
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 416 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-225-488-7

Query Match 3.2%; Score 7; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 139 LPAGIER 145
Db 75 LPAGIER 81
RESULT 2
US-08-769-309A-21
; Sequence 21, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-769-309A-21

Query Match 2.7%; Score 6; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 KTSFRK 20
Db 15 KTSFRK 20

RESULT 3
US-08-449-639-3
; Sequence 3, Application US/08449639
; Patent No. 5770191
; GENERAL INFORMATION:
; APPLICANT: Johnson, Howard M.
; APPLICANT: Szente, Brian E.
; TITLE OF INVENTION: C-terminal Peptides of Interferon-Gamma
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville
STATE: FL USA
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,639
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF150
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-375-8100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-449-639-3

Query Match 2.7%; Score 6; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 KHRVER 25
Db 24 KHRVER 29

RESULT 4
US-08-902-623-43
Sequence 43, Application US/08902623
Patent No. 5922545
GENERAL INFORMATION:
APPLICANT: MATTHEAKIS, LARRY C.
APPLICANT: DOWER, WILLIAM J.
TITLE OF INVENTION: IN VITRO PEPTIDE AND ANTIBODY DISPLAY
LIBRARIES
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/902,623
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/586,176
FILING DATE: 17-JAN-1996
APPLICATION NUMBER: US 08/300,262
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/144,775
FILING DATE: 29-OCT-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/12206
FILING DATE: US 25-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: DUNN, TRACY J.
REGISTRATION NUMBER: 34,587
REFERENCE/DOCKET NUMBER: 16528X-003230US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
US-08-902-623-43

Query Match 2.7%; Score 6; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 ROATRR 34
Db 83 ROATRR 88

RESULT 5
US-08-215-805A-82
Sequence 82, Application US/08215805A
Patent No. 5559008
GENERAL INFORMATION:
APPLICANT: Chang, Yung-Fu
TITLE OF INVENTION: LEUKOTOXIN GENE FROM PASTEURILLA
SUIS
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215,805A
FILING DATE: 22-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Imlan, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 19603/61 (D-1329A)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 165 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pasteurella suis
STRAIN: 5943
IMMEDIATE SOURCE:
LIBRARY: P. suis DNA in Bacteriophage lambda-dash
CLONE: (Lambda)yfc33-37

US-08-215-805A-82

Query Match 2.7%; Score 6; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 KRTGKI 110
|||||
DB 124 KRTGKI 129

RESULT 6
US-08-933-750C-28
; Sequence 28, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 179 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNCBOT03
; CLONE: 1577239
; US-08-933-750C-28

Query Match 2.7%; Score 6; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 TLPAGI 143
|||||
DB 126 TLPAGI 131

RESULT 7

US-08-622-352A-3
; Sequence 3, Application US/08622352A
; Patent No. 5824546
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Demaio, James
; TITLE OF INVENTION: REGULATION OF A SIGMA FACTOR
; TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,352A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; US-08-622-352A-3

Query Match 2.7%; Score 6; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 AAGRDL 81
|||||
DB 27 AAGRDL 32

RESULT 8

US-08-769-309A-3
; Sequence 3, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauer, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/769,309A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5741890and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 292 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-769-309A-3

Query Match 2.7%; Score 6; DB 1; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 KTSFRK 20
| | | | |
Db 45 KTSFRK 50

RESULT 9
US-08-375-962B-12
Sequence 12, Application US/08375962B
Patent No. 5731195
GENERAL INFORMATION:
APPLICANT: SIMON, ANDRAS; HELLMAN, ULF; WERNSTEDT,
APPLICANT: CHRISTER, ERIKSSON, ULF.
TITLE OF INVENTION: Isolated Nucleic Acid Molecule
TITLE OF INVENTION: Which Codes for A 32 kDa Protein Having 11-CIS Retinol
TITLE OF INVENTION: Dehydrogenase Activity, and Which Associates With P63,
TITLE OF INVENTION: A Portion of a Retinol Binding Protein Receptor
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect (ASCII standard)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/375,962B
FILING DATE: 20-January-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/258,418
FILING DATE: 6-October-1994
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: LUD 5372.1 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 688-3884
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

FEATURE:
NAME/KEY: Human estradiol 17-b dehydrogenase (hEDH)
US-08-375-962B-12

Query Match 2.7%; Score 6; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 DEAGRS 63
| | | | |
Db 307 DEAGRS 312

RESULT 10
US-08-562-114B-12
Sequence 12, Application US/08562114B
Patent No. 5972646
GENERAL INFORMATION:
APPLICANT: ERIKSSON ET AL.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE WHICH CODES FOR A
TITLE OF INVENTION: 32 KDA PROTEIN HAVING 11-CIS RETINOL HYDROGENASE
TITLE OF INVENTION: ACTIVITY, AND WHICH ASSOCIATES WITH P63, A
TITLE OF INVENTION: PORTION OF A RETINOL BINDING PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect 5.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/562,114B
FILING DATE: 22-No. 5972646ember-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/375,962
FILING DATE: 20-January-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kohli, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5372.2 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Human estradiol 17-b dehydrogenase (hEDH)
US-08-562-114B-12

Query Match 2.7%; Score 6; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 DEAGRS 63
| | | | |
Db 307 DEAGRS 312

RESULT 11
US-08-855-714-3
Sequence 3, Application US/08855714

Patent No. 5939075
GENERAL INFORMATION:
APPLICANT: HOUNG, HUO-SHU H.
TITLE OF INVENTION: MUTANTS OF BRUCELLA MELITENSIS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: John Moran, Esq.
STREET: HQ USAMRDC, Dept. of Army, Fort Detrick
CITY: Frederick
STATE: MD
COUNTRY: US
ZIP: 21702-5012
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,714
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/334,129
FILING DATE: 04-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: 08/143,692
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 619-2065
TELEFAX: (301) 619-7714
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE:
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BRUCELLA MELITENSIS
US-08-855-714-3

Query Match 2.7%; Score 6; DB 2; Length 339;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PTSWLK 171
Db 64 PTSWLK 69

RESULT 12
US-08-467-559B-9
Sequence 9, Application US/08467559B
Patent No. 5928890
GENERAL INFORMATION:
APPLICANT: LI, YI
TITLE OF INVENTION: HUMAN AMINE RECEPTOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: UNITED STATES OF AMERICA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,559B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0840000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-467-559B-9

Query Match 2.7%; Score 6; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 DSLANG 212
Db 259 DSLANG 264

RESULT 13
US-08-846-762-89
Sequence 89, Application US/08846762A
Patent No. 5994072
GENERAL INFORMATION:
APPLICANT: Lam, Joseph S.
APPLICANT: Burrows, Lori
APPLICANT: Charter, Deborah
APPLICANT: de Kievit, Teresa
TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
FILE REFERENCE: 6580-089
CURRENT APPLICATION NUMBER: US/08/846,762A
CURRENT FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 89
LENGTH: 403
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-08-846-762-89

Query Match 2.7%; Score 6; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 PYDVL 54
Db 98 PYDVL 103

RESULT 14
US-09-031-485-15
Sequence 15, Application US/09031485
Patent No. 5824306
GENERAL INFORMATION:
APPLICANT: Tang, Liang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

```

; TITLE OF INVENTION:  USES THEREOF
; NUMBER OF SEQUENCES:  85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Carol Talkington Verser, Ph.D.
; ADDRESSEE:  Heska Corporation
; STREET:  1825 Sharp Point Drive
; CITY:  Fort Collins
; STATE:  Colorado
; COUNTRY:  USA
; ZIP:  80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  Windows 95
; SOFTWARE:  WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/09/031,485
; FILING DATE:
; CLASSIFICATION:  530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US/08/847,429
; FILING DATE:  24-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME:  Verser, Carol Talkington
; REGISTRATION NUMBER:  37,459
; REFERENCE/DOCKET NUMBER:  HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  970/493-7272
; TELEFAX:  970/484-9505
; INFORMATION FOR SEQ ID NO:  15:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  409 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  protein
; US-09-031-485-15

Query Match      2.7%; Score 6; DB 2; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches  6; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

QY  38 PQPTRV 43
DB  210 PQPTRV 215

Search completed: June 10, 2000, 11:31:29
Job time: 343 sec

; TITLE OF INVENTION:  USES THEREOF
; NUMBER OF SEQUENCES:  85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Carol Talkington Verser, Ph.D.
; ADDRESSEE:  Heska Corporation
; STREET:  1825 Sharp Point Drive
; CITY:  Fort Collins
; STATE:  Colorado
; COUNTRY:  USA
; ZIP:  80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  Windows 95
; SOFTWARE:  WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/09/031,485
; FILING DATE:
; CLASSIFICATION:  530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US/08/847,429
; FILING DATE:  24-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME:  Verser, Carol Talkington
; REGISTRATION NUMBER:  37,459
; REFERENCE/DOCKET NUMBER:  HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  970/493-7272
; TELEFAX:  970/484-9505
; INFORMATION FOR SEQ ID NO:  15:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  409 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  protein
; US-09-031-485-15

Query Match      2.7%; Score 6; DB 2; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches  6; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

QY  38 PQPTRV 43
DB  210 PQPTRV 215

RESULT 15
US-08-847-429A-15
; Sequence 15, Application US/08847429A
; Patent No. 5827692
; GENERAL INFORMATION:
; APPLICANT:  Tang, Liang
; APPLICANT:  Blehm, E. Scot
; TITLE OF INVENTION:  DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION:  PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION:  USES THEREOF
; NUMBER OF SEQUENCES:  85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Carol Talkington Verser, Ph.D.
; ADDRESSEE:  Heska Corporation
; STREET:  1825 Sharp Point Drive
; CITY:  Fort Collins
; STATE:  Colorado
; COUNTRY:  USA
; ZIP:  80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  Windows 95
; SOFTWARE:  WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/847,429A
```

```

; FILING DATE:  24-APR-1997
; CLASSIFICATION:  435
; ATTORNEY/AGENT INFORMATION:
; NAME:  Verser, Carol Talkington
; REGISTRATION NUMBER:  37,459
; REFERENCE/DOCKET NUMBER:  HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  970/493-7272
; TELEFAX:  970/484-9505
; INFORMATION FOR SEQ ID NO:  15:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  409 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  protein
; US-08-847-429A-15

Query Match      2.7%; Score 6; DB 2; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches  6; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

QY  38 PQPTRV 43
DB  210 PQPTRV 215

Search completed: June 10, 2000, 11:31:29
Job time: 343 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 10:45:07 ; Search time 14.58 seconds

(without alignments)

892.716 Million cell updates/sec

Title: US-09-252-691-7056

Perfect score: 222

Sequence: 1 AIRMQLTPNTWTKTSFRK.....SYTLDSLANGWRDVTPEKN 222

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 168808 seqs, 58629743 residues

Word size : 0

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database : PIR_63.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	23.4	207	2 D64858	probable pseudouri
2	15	6.8	240	2 I64156	probable pseudouri
3	12	5.4	185	2 S74428	probable pseudouri
4	10	4.5	229	2 S45555	conserved hypothet
5	10	4.5	257	2 D75462	ribosomal large su
6	8	3.6	341	2 T13185	hypothetical prote
7	8	3.6	448	2 A71347	hypothetical prote
8	8	3.6	555	2 B41492	58K antigen - Rick
9	8	3.6	577	2 B75621	sensor histidine k
10	7	3.2	241	2 JE0169	dnaj heat shock pr
11	7	3.2	250	2 A43662	replicating protel
12	7	3.2	288	2 S41382	pobR protein - pse
13	7	3.2	312	2 S74669	hypothetical prote
14	7	3.2	406	2 A71965	hypothetical prote
15	7	3.2	444	2 T35794	probable 4-aminobu
16	7	3.2	514	2 B64634	hypothetical prote
17	7	3.2	514	2 E71880	probable outer mem
18	7	3.2	515	2 S75942	protein disulfide-
19	7	3.2	532	1 QFFPM	neurofilament trip
20	7	3.2	534	2 S56785	TCPI-related prote
21	7	3.2	546	2 S48086	t-complex-type mol
22	7	3.2	604	2 S58643	probable transcrip
23	7	3.2	707	1 A53796	gelatinase B (EC 3
24	7	3.2	764	2 S49849	aconitate hydratase
25	7	3.2	845	2 A45669	neurofilament trip
26	7	3.2	849	2 S00030	neurofilament trip
27	7	3.2	898	2 T10101	aconitate hydratase
28	7	3.2	916	2 A27864	neurofilament trip
29	7	3.2	990	1 TQEC26	transposase - Esch

probable chitin sy
transcription/repa
nuclear pore prote
cholic acid transp
hypothetical prote
hypothetical prote
hypothetical prote
10K cell fusion pr
prgs protein - Ent
probable pheromone
probable ROK fami
integration host f
hypothetical prote
colicin V - Escher
hypothetical prote
conserved hypothet
hypothetical prote
hypothetical prote
interleukin-4 - pi
histone H1 - Euplo
hypothetical prote
hypothetical prote
membrane-bound mul
probable maturase-
gene aac(6')-Ih pr
hypothetical prote
Huntington's disea
ureE protein - Kle
merozoite surface
endodeoxyribonucle
hypothetical prote
polyprotein - chic
gamma-D-crystallin
plu protein - frui
probable succinate
hypothetical prote
apolipoprotein 2 p
beta-glucosidase (
hypothetical prote
hypothetical prote
beta-crystallin B2
hypothetical prote
extragenic suppres
hypothetical prote
homotetic protein H
homotetic protein H
hypothetical prote
probable drought-1
probable mra prot
hypothetical prote
hypothetical 25.9
hypothetical prote
nitrate hydratase
probable glucose-6
hypothetical prote
tryptophan synthas
zinc ABC transport
probable rsbw prot
hypothetical prote
pseudouridyate sy
hypothetical prote
hypothetical prote
N-acetylmuramoyl-L
lmbj protein - Str
phosphomannomutase
hypothetical prote
probable ribD - My
lysine-arginine-or
hypothetical prote
hydrolyase - Strept
H+-transporting Ar

103	6	2.7	266	2	S06462	beta-lactamase (EC	176	6	2.7	377	2	D70957	hypothetical prote
104	6	2.7	266	2	S33163	beta-lactamase (EC	177	6	2.7	377	2	T00643	zinc metalloprotei
105	6	2.7	269	2	T36910	hypothetical prote	178	6	2.7	377	2	B71184	probable sarcosine
106	6	2.7	270	2	T22151	hypothetical prote	179	6	2.7	379	2	F75144	sarcosine oxidase,
107	6	2.7	271	2	A72200	septum site-determ	180	6	2.7	383	2	F72352	hypothetical prote
108	6	2.7	272	1	A26936	NAD+ synthase (EC	181	6	2.7	383	2	G72777	probable S2P metal
109	6	2.7	274	2	H70900	hypothetical prote	182	6	2.7	386	2	S41497	thyroid hormone re
110	6	2.7	276	2	T03778	hypothetical prote	183	6	2.7	386	2	T35854	probable ABC trans
111	6	2.7	277	2	S38503	beta-lactamase (EC	184	6	2.7	388	2	H64427	hypothetical prote,
112	6	2.7	281	2	S38671	beta-lactamase (EC	185	6	2.7	388	2	C70754	probable regulator
113	6	2.7	281	2	E72395	sugar ABC transpor	186	6	2.7	390	2	G72491	hypothetical prote
114	6	2.7	281	2	G71204	hypothetical prote	187	6	2.7	391	2	S74688	hypothetical prote
115	6	2.7	281	2	E75018	endonuclease iv re	188	6	2.7	391	2	T32429	hypothetical prote
116	6	2.7	284	2	S62931	probable membrane	189	6	2.7	391	3	T17486	cytochrome P450 hy
117	6	2.7	284	2	S62955	probable membrane	190	6	2.7	393	2	E72549	probable acyl-CoA
118	6	2.7	285	2	T14717	yhgA protein homol	191	6	2.7	395	2	G75587	hypothetical prote
119	6	2.7	288	2	S44256	fructokinase (EC 2	192	6	2.7	397	2	S45496	isp7 protein - fls
120	6	2.7	288	2	A71148	hypothetical prote	193	6	2.7	398	2	B70752	hypothetical prote
121	6	2.7	290	1	H70530	probable extrageni	194	6	2.7	399	1	A43685	polymerase-associ
122	6	2.7	290	2	S72970	tyrosine--trNA lig	195	6	2.7	402	2	S32565	actin-binding prot
123	6	2.7	294	2	T35338	probable transcrip	196	6	2.7	403	1	H64142	hypothetical prote
124	6	2.7	298	2	T18046	probable DNA ligas	197	6	2.7	403	2	S70576	glycosyltransferas
125	6	2.7	298	2	B75612	phosphate ABC tran	198	6	2.7	403	2	T34523	hypothetical ROK f
126	6	2.7	300	1	H70344	6-phosphogluconate	199	6	2.7	403	2	A48447	hypothetical associat
127	6	2.7	300	2	T13535	N-acetylmuramoyl-L	200	6	2.7	406	3	T17487	cytochrome P450 hy
128	6	2.7	301	2	A43355	CD44 glycoprotein	201	6	2.7	407	2	F69805	hypothetical prote
129	6	2.7	303	2	A75301	SudB-related prote	202	6	2.7	407	2	F70696	hypothetical prote
130	6	2.7	306	2	A48118	major epidermal ca	203	6	2.7	407	2	T39326	pseudouridine synt
131	6	2.7	307	2	T34651	probable transmemb	204	6	2.7	408	2	G69874	hypothetical prote
132	6	2.7	312	2	E69045	8-oxoguanine DNA g	205	6	2.7	413	2	B44285	methyloaspartate am
133	6	2.7	313	1	XWNC	ADP, ATP carrier pr	206	6	2.7	413	2	T26572	hypothetical prote
134	6	2.7	313	2	S41385	hypothetical yadB	207	6	2.7	414	1	JQ1016	glycine hydroxymet
135	6	2.7	313	2	T35177	iron-sulfur oxidor	208	6	2.7	414	2	S30397	CD44 protein - mou
136	6	2.7	318	1	D84417	transformation sen	209	6	2.7	415	1	K1PPPG	phosphoglycerate k
137	6	2.7	319	2	S61293	transcription init	210	6	2.7	416	2	G64647	hypothetical prote
138	6	2.7	319	2	A69756	adhesion protein h	211	6	2.7	416	2	D71936	hypothetical prote
139	6	2.7	321	2	H65173	hypothetical trans	212	6	2.7	425	2	T04743	hypothetical prote
140	6	2.7	321	2	C70653	probable prephenat	213	6	2.7	428	2	T21989	cyclin cycli - Arab
141	6	2.7	323	2	A36444	rho-crystallin - b	214	6	2.7	432	2	I56934	cytotoxic T-lympho
142	6	2.7	323	2	F75532	MutT/nudix family	215	6	2.7	432	2	A27447	fibrinogen-like pr
143	6	2.7	324	2	JC4280	rho-crystallin - b	216	6	2.7	433	2	T21989	hypothetical prote
144	6	2.7	324	2	T27302	hypothetical prote	217	6	2.7	436	2	D72472	hypothetical prote
145	6	2.7	328	1	DEHUE7	probable transposa	218	6	2.7	438	2	T04655	probable electron
146	6	2.7	328	2	T14989	hypothetical prote	219	6	2.7	439	2	I37391	protein kinase hom
147	6	2.7	328	2	T19815	hypothetical prote	220	6	2.7	444	2	E69102	fibrinogen-like pr
148	6	2.7	328	2	S58158	estradiol 17beta-d	221	6	2.7	445	2	S77010	sensory transducti
149	6	2.7	329	2	T38562	hypothetical prote	222	6	2.7	447	2	B64894	nickel resistance
150	6	2.7	329	2	T38562	hypothetical prote	223	6	2.7	447	2	S52391	hypothetical prote
151	6	2.7	330	2	S46819	probable histone p	224	6	2.7	449	2	B75451	centrosomin B - mo
152	6	2.7	331	2	T15458	pseudouridine synt	225	6	2.7	450	2	H64741	NADH oxidase - Del
153	6	2.7	338	2	C71603	probable pseudouri	226	6	2.7	450	2	G70210	yael protein - Esc
154	6	2.7	338	3	T40327	hypothetical prote	227	6	2.7	452	2	C70728	conserved hypothet
155	6	2.7	340	2	G70741	conserved hypothet	228	6	2.7	456	2	T32300	hypothetical prote
156	6	2.7	342	2	C71372	hypothetical prote	229	6	2.7	457	2	D69962	hypothetical prote
157	6	2.7	344	2	T21604	holocytochrome-c s	230	6	2.7	461	2	B72625	dihydroliipoamide d
158	6	2.7	346	2	A34365	probable peroxidase	231	6	2.7	464	2	S12591	probable mercuric
159	6	2.7	351	2	T09163	Li protein - rhesus	232	6	2.7	465	2	I51661	beta-1-adrenergic
160	6	2.7	354	1	PLWL81	recf protein - Sal	233	6	2.7	466	2	S36794	beta-1-adrenergic
161	6	2.7	355	2	S21057	SMR1 protein - pod	234	6	2.7	474	2	H70896	probable fumarate
162	6	2.7	356	2	S39889	hypothetical prote	235	6	2.7	485	2	T27147	hypothetical prote
163	6	2.7	357	1	RQECF	recf protein - Esc	236	6	2.7	488	1	I39769	aldehyde dehydroge
164	6	2.7	359	2	S27788	neutrophil oxidase	237	6	2.7	490	2	T36716	probable penicilli
165	6	2.7	360	1	A25732	inhibin alpha chal	238	6	2.7	490	2	B71279	probable serine-ty
166	6	2.7	362	1	E70892	recf protein - Pro	239	6	2.7	498	2	T34592	DEAD-box RNA helic
167	6	2.7	363	2	J00735	mannopine biosynth	240	6	2.7	499	1	PLWL13	L1 protein - human
168	6	2.7	363	2	I39726	NS5 protein - hepa	241	6	2.7	499	1	PLWL33	L1 protein - human
169	6	2.7	368	1	JQ0880	X-Pro dipeptidase	242	6	2.7	500	1	PLWL6	L1 protein - human
170	6	2.7	371	2	S52202	mitogen-activated	243	6	2.7	501	1	A40938	cytochrome P450 ib
171	6	2.7	371	2	S60121	flagellar motor sw	244	6	2.7	501	1	PLWL11	L1 protein - human
172	6	2.7	371	2	H72045	recf protein - Hae	245	6	2.7	501	2	T04735	cytochrome P450 ho
173	6	2.7	375	1	I64106	hypothetical prote	246	6	2.7	501	2	JC5739	membrane-bound pro
174	6	2.7	375	2	A75464		247	6	2.7	502	1	PLWL1	L1 protein - pygmy
175	6	2.7	375	2			248	6	2.7	502	1		

249	6	2.7	502	1	PLWL42	L1 protein - human	322	6	2.7	649	2	T18063	DNA binding protei
250	6	2.7	503	2	S36549	L1 protein - human	323	6	2.7	651	2	JC4610	dnak-type molecula
251	6	2.7	504	1	PLWLP	L1 protein - human	324	6	2.7	652	2	S64073	hypothetical prote
252	6	2.7	505	1	PLWL39	L1 protein - human	325	6	2.7	657	2	D71351	probable prinosoma
253	6	2.7	506	2	JC5054	high affinity nitr	326	6	2.7	661	2	A53668	glycoprotein gp100
254	6	2.7	507	2	S36508	L1 protein - human	327	6	2.7	662	2	T38400	melanoma-associate
255	6	2.7	508	2	T31630	hypothetical prote	328	6	2.7	663	1	E69475	DNA topoisomerase
256	6	2.7	509	2	JC5055	high affinity nitr	329	6	2.7	665	2	A30185	nuclear protein nu
257	6	2.7	510	1	S13620	L1 protein - human	330	6	2.7	668	2	A41234	melanocyte-specifi
258	6	2.7	511	1	S15627	L1 protein - human	331	6	2.7	669	2	T38181	glutamate--cystein
259	6	2.7	512	2	T36306	probable anthranil	332	6	2.7	672	2	S61462	p83/100 protein -
260	6	2.7	513	2	T30924	hypothetical prote	333	6	2.7	672	2	S61463	p83/100 protein -
261	6	2.7	514	2	B72752	hypothetical prote	334	6	2.7	674	2	T19495	hypothetical prote
262	6	2.7	515	2	T35371	probable Na+/H+ an	335	6	2.7	677	1	S90978	chromogranin B pre
263	6	2.7	516	2	T35534	hypothetical prote	336	6	2.7	677	2	J50178	protein kinase YKR
264	6	2.7	517	1	CSBYP	catalase (EC 1.1.1.	337	6	2.7	679	2	E71282	probable NH(3)-dep
265	6	2.7	518	2	I49154	calcitonin recepto	338	6	2.7	684	2	S76658	hypothetical prote
266	6	2.7	519	2	T25615	hypothetical prote	339	6	2.7	685	2	JC6331	rho-type guanine e
267	6	2.7	520	2	T05625	hypothetical prote	340	6	2.7	692	2	T32980	hypothetical prote
268	6	2.7	521	1	S34190	sulfite reductase	341	6	2.7	693	2	S61464	p83/100 protein -
269	6	2.7	522	1	PLWL58	L1 protein - human	342	6	2.7	693	2	I40090	p93 protein - Lyme
270	6	2.7	523	2	D72107	oligopeptide bindi	343	6	2.7	693	2	I40141	p93 protein - Lyme
271	6	2.7	524	2	S36578	L1 protein - human	344	6	2.7	698	2	A59222	hypothetical prote
272	6	2.7	525	2	S36537	L1 protein - human	345	6	2.7	698	2	T32840	hypothetical prote
273	6	2.7	526	2	S36554	L1 protein (altern	346	6	2.7	709	2	T16176	hypothetical prote
274	6	2.7	527	2	S36554	L1 protein (altern	347	6	2.7	711	2	F75440	methionine--CoA
275	6	2.7	528	2	S36554	actVA-1 protein -	348	6	2.7	722	2	S52205	meso protein - Leu
276	6	2.7	529	2	S36583	L1 protein - human	349	6	2.7	722	2	T38220	hypothetical prote
277	6	2.7	530	2	S36566	secreted acid phos	350	6	2.7	736	2	T00343	hypothetical prote
278	6	2.7	531	2	S36566	L1 protein - human	351	6	2.7	746	2	T06017	subtilisin-like pr
279	6	2.7	532	2	S24389	hybrid cluster (4F	352	6	2.7	746	2	S27985	cellulase (EC 3.2.
280	6	2.7	533	2	F69376	hybrid cluster (4F	353	6	2.7	760	2	T01441	hypothetical prote
281	6	2.7	534	2	E64395	hybrid cluster (4F	354	6	2.7	775	1	J01638	outer layer protei
282	6	2.7	535	2	S56152	H+-transporting AT	355	6	2.7	775	1	VFXRWF	outer layer protei
283	6	2.7	536	2	C75557	hypothetical prote	356	6	2.7	779	2	S50054	chloride channel p
284	6	2.7	537	2	T25195	hypothetical prote	357	6	2.7	786	2	H59980	single-strand DNA-
285	6	2.7	538	1	S29861	hybrid cluster (4F	358	6	2.7	790	2	T35095	hypothetical prote
286	6	2.7	539	1	SRECO1	glutamine--trNA li	359	6	2.7	797	2	T05247	methionine--trNA l
287	6	2.7	540	2	T02445	probable small nuc	360	6	2.7	802	2	T21315	hypothetical prote
288	6	2.7	541	2	S67452	hypothetical prote	361	6	2.7	810	1	I40508	endopeptidase Clp
289	6	2.7	542	2	T38067	hypothetical prote	362	6	2.7	814	2	S51451	probable membrane
290	6	2.7	543	2	T41952	hypothetical prote	363	6	2.7	823	2	S48911	hypothetical prote
291	6	2.7	544	2	T11581	viron protein - hu	364	6	2.7	835	2	E72305	hypothetical prote
292	6	2.7	545	2	T38942	nuclear scaffold-1	365	6	2.7	840	2	E71357	hypothetical prote
293	6	2.7	546	2	PLWL18	heat shock protein	366	6	2.7	841	2	E71808	probable iron (III
294	6	2.7	547	1	A69511	L1 protein - human	367	6	2.7	842	2	H64694	iron(III) dicitrat
295	6	2.7	548	1	CSBYT	C conserved hypoth	368	6	2.7	847	2	A48228	beta-N-acetylhexos
296	6	2.7	549	2	T19880	catalase (EC 1.1.1.	369	6	2.7	856	2	G70483	pyruvate water dik
297	6	2.7	550	2	F64110	hypothetical prote	370	6	2.7	859	2	S49046	luxQ protein - Vib
298	6	2.7	551	2	D65073	single-stranded DN	371	6	2.7	862	2	T05941	lipoxigenase (EC 1
299	6	2.7	552	2	C36329	hypothetical prote	372	6	2.7	863	2	S37088	phycobilisome anch
300	6	2.7	553	2	I50419	s-glycerin precursor	373	6	2.7	864	2	T05945	lipoxigenase (EC 1
301	6	2.7	554	2	C70893	hypothetical prote	374	6	2.7	865	1	S23454	lipoxigenase (EC 1
302	6	2.7	555	2	F71141	hypothetical prote	375	6	2.7	869	1	A28443	phosphatidylethano
303	6	2.7	556	2	A04337	glutamic acid-rich	376	6	2.7	877	2	S72541	nitrate reductase
304	6	2.7	557	2	A45135	profilaggrin - hum	377	6	2.7	878	1	A40091	interleukin-3 rece
305	6	2.7	558	2	S36502	L1 protein (altern	378	6	2.7	879	2	S64755	hypothetical prote
306	6	2.7	559	2	T38241	probable U3 small	379	6	2.7	896	1	A35782	cytokine receptor
307	6	2.7	560	2	T12994	hypothetical prote	380	6	2.7	896	2	I56563	interleukin-3 rece
308	6	2.7	561	2	G711682	GTP-binding protei	381	6	2.7	897	1	A39255	cytokine receptor
309	6	2.7	562	2	A53583	geranylgeranyl-dip	382	6	2.7	906	2	S82517	hypothetical prote
310	6	2.7	563	2	T08180	PF20 protein, micr	383	6	2.7	906	2	T39123	hypothetical prote
311	6	2.7	564	2	T21747	hypothetical prote	384	6	2.7	910	2	S64155	hypothetical prote
312	6	2.7	565	1	ERADFM	fiber protein - mo	385	6	2.7	910	2	S38167	hypothetical prote
313	6	2.7	566	2	S38802	hypothetical prote	386	6	2.7	921	2	S39295	beta-adaptin 1 - f
314	6	2.7	567	2	T37257	hypothetical prote	387	6	2.7	946	2	T38100	p115 - human
315	6	2.7	568	2	S59234	glutamate--cystein	388	6	2.7	947	1	K1ZMPO	pyruvate,orthophos
316	6	2.7	569	2	S38871	fmel 17 protein%	389	6	2.7	947	2	T02979	pyruvate,orthophos
317	6	2.7	570	2	T14776	hypothetical prote	390	6	2.7	949	1	S55478	pyruvate,orthophos
318	6	2.7	571	2	A45949	merozoite surface	391	6	2.7	951	2	D75377	probable proteinase
319	6	2.7	572	2	F70444	threonine--trNA li	392	6	2.7	953	2	S12894	pyruvate,orthophos
320	6	2.7	573	2	S67656	hypothetical prote	393	6	2.7	953	2	S56650	pyruvate,orthophos
321	6	2.7	574	1	A55093	fatty acid transpo	394	6	2.7	955	2	S56649	pyruvate,orthophos

541	5	2.3	65	2	T28602	hypothetical prote	614	5	2.3	94	2	S44903	zk552.3 protein -
542	5	2.3	65	2	T15856	hypothetical prote	615	5	2.3	94	3	T37723	hypothetical prote
543	5	2.3	65	2	A72172	H2L protein - vari	616	5	2.3	95	1	E64043	citrate (pro-3S)-1
544	5	2.3	68	2	JW0050	G-protein gamma 7	617	5	2.3	95	2	F64709	hypothetical prote
545	5	2.3	68	2	A45128	GTP-binding regula	618	5	2.3	95	2	T26179	hypothetical prote
546	5	2.3	68	2	T27270	hypothetical prote	619	5	2.3	96	2	S65794	glutathione peroxi
547	5	2.3	68	2	T29417	hypothetical prote	620	5	2.3	96	2	A05308	tissue kallikrein
548	5	2.3	69	2	I36580	G-protein gamma su	621	5	2.3	96	2	T07822	cystein proteinase
549	5	2.3	69	2	A71330	hypothetical prote	622	5	2.3	96	2	B46018	apolipoprotein C-I
550	5	2.3	70	1	B64091	ribosomal protein	623	5	2.3	96	2	JN0572	neutrophil chemo-a
551	5	2.3	70	2	S73211	ribosomal protein	624	5	2.3	96	2	S00119	hypothetical prote
552	5	2.3	70	2	T28640	Y4JM protein - Rhi	625	5	2.3	96	2	C72309	hypothetical prote
553	5	2.3	70	2	F69058	hypothetical prote	626	5	2.3	97	1	FEBQ	ferredoxin [2Fe-2S
554	5	2.3	70	2	A05211	hypothetical prote	627	5	2.3	97	2	S69935	ferredoxin [2Fe-2S
555	5	2.3	71	2	C75627	hypothetical prote	628	5	2.3	97	2	B45271	small hydrophobic
556	5	2.3	71	2	T15836	hypothetical prote	629	5	2.3	97	2	T34765	keratin, feather (
557	5	2.3	72	2	S78251	ribosomal protein	630	5	2.3	98	2	S06808	hypothetical prote
558	5	2.3	73	2	S77494	ribosomal protein	631	5	2.3	98	2	S14883	hypothetical prote
559	5	2.3	74	1	H5Y34	heat shock protein	632	5	2.3	99	2	F64317	hypothetical prote
560	5	2.3	74	2	F70526	hypothetical prote	633	5	2.3	99	2	C71198	hypothetical prote
561	5	2.3	75	2	T12080	low molecularweig	634	5	2.3	99	2	T34614	NADH dehydrogenase
562	5	2.3	75	2	T01886	hypothetical prote	635	5	2.3	99	2	S22957	hypothetical prote
563	5	2.3	75	2	T36474	hypothetical prote	636	5	2.3	99	2	T39208	very hypothetical
564	5	2.3	76	2	A46171	hypothetical prote	637	5	2.3	100	1	VKLJG2	trans-regulatory s
565	5	2.3	76	2	B65076	hypothetical prote	638	5	2.3	100	1	VKLJCA	trans-regulatory s
566	5	2.3	77	2	S30988	gene 43 protein -	639	5	2.3	100	2	S34400	NADH dehydrogenase
567	5	2.3	77	2	T02773	Y4DJ protein - Rhi	640	5	2.3	100	2	S48180	mabinlin IV - Yun
568	5	2.3	78	2	T30471	hypothetical prote	641	5	2.3	100	2	S33097	rev protein - huma
569	5	2.3	78	2	J01781	SalF6R protein - v	642	5	2.3	100	2	E59460	conserved hypothet
570	5	2.3	79	2	F75343	hypothetical prote	643	5	2.3	100	2	S72816	hypothetical prote
571	5	2.3	79	2	G89294	conserved hypothet	644	5	2.3	100	2	G59268	hypothetical prote
572	5	2.3	79	2	T03721	reverse transcript	645	5	2.3	100	2	C72635	hypothetical prote
573	5	2.3	80	1	OSHU7B	cytochrome-c oxida	646	5	2.3	100	4	S32195	hypothetical prote
574	5	2.3	80	2	A48284	intestinal trefoll	647	5	2.3	101	1	MNVUMB	nonstructural prot
575	5	2.3	80	2	J01959	inorganic carbon t	648	5	2.3	101	1	MNVUBV	nonstructural prot
576	5	2.3	81	1	W2BE49	gene 49 protein -	649	5	2.3	101	2	B47624	Ig heavy chain V-I
577	5	2.3	81	2	T36197	probable acyl carr	650	5	2.3	101	2	G71017	hypothetical prote
578	5	2.3	81	2	T26500	hypothetical prote	651	5	2.3	102	2	S78181	NADH dehydrogenase
579	5	2.3	82	1	BVEC2E	cell growth regula	652	5	2.3	102	2	C41846	recombination prot
580	5	2.3	82	2	D60396	antigen 7H8/7 - p1	653	5	2.3	102	2	D41846	recombination prot
581	5	2.3	82	2	E69453	hypothetical prote	654	5	2.3	102	2	JB0036	hypothetical iss p
582	5	2.3	83	2	T17809	hypothetical prote	655	5	2.3	102	2	S00739	hypothetical prote
583	5	2.3	83	2	T42102	hypothetical prote	656	5	2.3	102	2	A34862	probable adhesion
584	5	2.3	83	2	G70765	hypothetical prote	657	5	2.3	102	2	T39580	uvr3lp - fission y
585	5	2.3	84	2	S78266	hypothetical prote	658	5	2.3	102	2	T16931	hypothetical prote
586	5	2.3	84	2	T36353	probable ribosomal	659	5	2.3	102	2	E72479	hypothetical prote
587	5	2.3	85	2	A28561	late lactation pro	660	5	2.3	103	1	VKLJGG	trans-regulatory s
588	5	2.3	87	2	C49917	probable pyruvate,	661	5	2.3	103	2	S12158	rev protein - huma
589	5	2.3	87	2	B46200	retrovirus-related	662	5	2.3	103	2	T05290	hypothetical prote
590	5	2.3	87	2	S34444	pol polyprotein -	663	5	2.3	103	2	B34862	probable adhesion
591	5	2.3	87	2	S32445	pol polyprotein -	664	5	2.3	103	2	B71053	hypothetical prote
592	5	2.3	87	2	S32447	pol polyprotein -	665	5	2.3	103	2	A71071	hypothetical prote
593	5	2.3	87	2	S32449	pol polyprotein -	666	5	2.3	104	2	S15395	tissue kallikrein-
594	5	2.3	87	2	S60786	M protein precurs	667	5	2.3	104	2	A70665	probable ureb prot
595	5	2.3	87	2	T22931	hypothetical prote	668	5	2.3	104	2	S48178	mabinlin III - Yun
596	5	2.3	87	2	A35666	transcription acti	669	5	2.3	104	2	A42721	recombination prot
597	5	2.3	87	2	T00171	hypothetical prote	670	5	2.3	104	2	B42721	recombination prot
598	5	2.3	88	1	OSB07B	cytochrome-c oxida	671	5	2.3	104	2	C42721	recombination prot
599	5	2.3	88	2	T36927	probable small sec	672	5	2.3	104	2	E42721	recombination prot
600	5	2.3	89	2	D71269	probable ribosomal	673	5	2.3	104	2	F42721	recombination prot
601	5	2.3	89	2	E75101	hypothetical prote	674	5	2.3	104	2	D42721	recombination prot
602	5	2.3	90	2	E70694	hypothetical prote	675	5	2.3	104	2	S43805	hypothetical prote
603	5	2.3	90	2	S72290	ribosomal protein	676	5	2.3	104	2	A75348	conserved hypothet
604	5	2.3	91	1	A46539	monocyte chemoatr	677	5	2.3	104	2	T41245	replication factor
605	5	2.3	91	2	E70213	hypothetical prote	678	5	2.3	104	2	T29965	hypothetical prote
606	5	2.3	91	2	T03202	hypothetical prote	679	5	2.3	104	2	H72553	hypothetical prote
607	5	2.3	91	2	T22335	hypothetical prote	680	5	2.3	105	2	S76849	hypothetical prote
608	5	2.3	92	2	S41662	L-phenylalanine ox	681	5	2.3	105	2	F72658	hypothetical prote
609	5	2.3	93	2	S61391	hypothetical prote	682	5	2.3	106	2	H71242	hypothetical prote
610	5	2.3	93	2	T35668	nitrogen regulator	683	5	2.3	107	1	VKLJST	trans-regulatory s
611	5	2.3	94	1	VUWTEM	embryonic abundan	684	5	2.3	107	2	S23472	cpcF protein - Fis
612	5	2.3	94	2	A64111	integration host f	685	5	2.3	107	2	E72710	hypothetical prote
613	5	2.3	94	2	I40210	probable ferredoxi	686	5	2.3	108	1	H64913	probable membrane

687	5	2.3	108	1	S73448	hypothetical prote	760	5	2.3	120	2	D69696	ribosomal protein
688	5	2.3	108	2	S40149	integrin alpha-7C	761	5	2.3	120	2	F75171	lsu ribosomal prot
689	5	2.3	108	2	S09863	hypothetical prote	762	5	2.3	120	2	B42573	urf120 - Paracoccu
690	5	2.3	108	2	T06898	hypothetical prote	763	5	2.3	120	2	A70799	probable transcrip
691	5	2.3	108	2	T38936	non-histone chromo	764	5	2.3	120	2	S29267	protein kinase C s
692	5	2.3	108	2	S65757	hypothetical prote	765	5	2.3	120	3	T39807	hypothetical prote
693	5	2.3	109	2	T03965	allergenic protein	766	5	2.3	121	2	D32268	carcinoembryonic a
694	5	2.3	109	2	S29655	juvenile-hormone e	767	5	2.3	121	2	F71086	hypothetical prote
695	5	2.3	109	2	A72622	ubiquinone-binding	768	5	2.3	121	2	F44221	orf6 protein - Aut
696	5	2.3	110	1	UYBO	T-cell receptor al	769	5	2.3	122	2	C70971	hypothetical prote
697	5	2.3	110	2	S22890	probable sigma reg	770	5	2.3	122	2	C70983	hypothetical prote
698	5	2.3	110	2	H71472	hypothetical prote	771	5	2.3	122	2	T29631	hypothetical prote
699	5	2.3	110	2	T12730	hypothetical prote	772	5	2.3	123	2	D41870	probable membrane
700	5	2.3	110	2	T08921	hypothetical prote	773	5	2.3	123	2	S70044	guanine-nucleotide
701	5	2.3	110	2	A38529	nika protein - Esc	774	5	2.3	123	2	I52427	glucagon 1 precurs
702	5	2.3	111	2	D70899	ubiquinone-binding	775	5	2.3	123	2	S29714	glucagon 1 precurs
703	5	2.3	111	2	A32450	T cell receptor Ex	776	5	2.3	124	1	GCAF	hypothetical prote
704	5	2.3	111	2	S57889	probable L23 ribos	777	5	2.3	124	2	G70785	hypothetical prote
705	5	2.3	111	2	D71507	nifu protein homol	778	5	2.3	124	2	D71100	hypothetical prote
706	5	2.3	111	2	C70024	conserved hypothet	779	5	2.3	124	2	T39259	hypothetical prote
707	5	2.3	111	2	B69168	hypothetical prote	780	5	2.3	124	2	S71888	hypothetical prote
708	5	2.3	111	2	B75605	probable electron	781	5	2.3	124	2	T38491	S-phase inhibitor
709	5	2.3	111	2	T36016	hypothetical prote	782	5	2.3	125	2	S41139	Spdip - fission ye
710	5	2.3	111	2	C72714	hypothetical prote	783	5	2.3	125	2	B70307	ribosomal protein
711	5	2.3	112	2	SI4355	glutathione transf	784	5	2.3	125	2	T11941	ribosomal protein
712	5	2.3	112	2	G70747	probable nitrogen	785	5	2.3	125	2	T03147	ribosomal protein
713	5	2.3	112	2	D69543	hypothetical prote	786	5	2.3	125	2	T05193	hypothetical prote
714	5	2.3	112	2	F75628	hypothetical prote	787	5	2.3	125	2	A53627	hypothetical prote
715	5	2.3	112	4	S62170	hypothetical prote	788	5	2.3	125	2	B72521	fibroblast growth
716	5	2.3	113	2	T09627	positive transcrip	789	5	2.3	126	2	PC2257	hypothetical prote
717	5	2.3	113	2	C70691	hypothetical prote	790	5	2.3	126	2	A38154	cytochrome P450 pr
718	5	2.3	114	2	S00996	hypothetical prote	791	5	2.3	126	2	S32906	oogenesis required
719	5	2.3	114	2	A55010	neutrophil-activat	792	5	2.3	126	2	T09789	hypothetical prote
720	5	2.3	114	2	F69782	hypothetical prote	793	5	2.3	126	2	D71239	abscisic acid- and
721	5	2.3	115	2	S39393	glutathione transf	794	5	2.3	127	1	VS8H2	hypothetical prote
722	5	2.3	115	2	T28769	hypothetical prote	795	5	2.3	127	2	T03866	purithionin II pre
723	5	2.3	116	2	S44709	opacity protein op	796	5	2.3	127	2	S42606	hypothetical prote
724	5	2.3	116	2	PC4092	alanine carrier pr	797	5	2.3	127	2	A71538	hypothetical prote
725	5	2.3	116	2	S71433	hypothetical prote	798	5	2.3	128	2	C06985	probable phenylala
726	5	2.3	116	2	T14876	hypothetical prote	799	5	2.3	128	2	T72804	probable deoxycyti
727	5	2.3	116	2	E70510	hypothetical prote	800	5	2.3	128	2	A75540	hypothetical prote
728	5	2.3	116	2	B64451	hypothetical prote	801	5	2.3	128	2	E64463	hypothetical prote
729	5	2.3	116	2	B69270	hypothetical prote	802	5	2.3	128	2	F69310	conserved hypothet
730	5	2.3	116	2	T05517	abscisic acid-indu	803	5	2.3	128	2	D72750	hypothetical prote
731	5	2.3	116	2	D72471	hypothetical prote	804	5	2.3	129	2	S03534	Ig heavy chain pre
732	5	2.3	117	2	S21668	Ig kappa chain V r	805	5	2.3	129	2	S34553	ribosomal protein
733	5	2.3	117	2	D71983	aspartate 1-decarb	806	5	2.3	129	2	I61187	alpha-7A integrin
734	5	2.3	117	2	B64524	aspartate 1-decarb	807	5	2.3	129	2	T03047	hypothetical prote
735	5	2.3	117	2	T39528	very hypothetical	808	5	2.3	129	2	D70153	hypothetical prote
736	5	2.3	117	2	F72869	hypothetical prote	809	5	2.3	129	2	A69139	hypothetical prote
737	5	2.3	117	2	B72508	hypothetical prote	810	5	2.3	130	1	R31V11	ribosomal protein
738	5	2.3	118	1	S30566	beta-2-microglobul	811	5	2.3	130	2	S77479	ribosomal protein
739	5	2.3	118	2	G71833	ribosomal protein	812	5	2.3	130	2	T07518	ribosomal protein
740	5	2.3	118	2	T09487	hypothetical prote	813	5	2.3	130	2	C58546	flagellar protein
741	5	2.3	118	2	S75220	hypothetical prote	814	5	2.3	130	2	T08532	traH protein - Ent
742	5	2.3	118	2	B69047	conserved hypothet	815	5	2.3	130	2	S22993	traH protein - Esc
743	5	2.3	118	2	T10375	apoptosis inhibito	816	5	2.3	130	2	F70908	hypothetical prote
744	5	2.3	118	2	H72677	hypothetical prote	817	5	2.3	130	2	C70811	hypothetical prote
745	5	2.3	118	2	A72474	hypothetical prote	818	5	2.3	130	2	T12478	hypothetical prote
746	5	2.3	119	2	T14396	lipid transfer pro	819	5	2.3	131	1	R38851	ribosomal protein
747	5	2.3	119	2	B27588	Ig kappa chain pre	820	5	2.3	131	2	S23511	thionin type v pre
748	5	2.3	119	2	A27588	Ig kappa chain pre	821	5	2.3	131	2	E29774	T-cell receptor al
749	5	2.3	119	2	G64682	ribosomal protein	822	5	2.3	131	2	S75682	hypothetical prote
750	5	2.3	119	2	S44711	opacity protein op	823	5	2.3	131	2	H75323	hypothetical prote
751	5	2.3	119	2	S54621	probable membrane	824	5	2.3	131	2	T26286	hypothetical prote
752	5	2.3	119	2	E75319	ferredoxin - Deino	825	5	2.3	131	2	T34088	spermiogenesis pro
753	5	2.3	119	2	T35859	hypothetical prote	826	5	2.3	132	1	MTONIK	melanin-concentrat
754	5	2.3	119	2	T36326	probable dihydron	827	5	2.3	132	1	MTON2K	melanin-concentrat
755	5	2.3	119	2	T28765	hypothetical prote	828	5	2.3	132	2	B32910	melanin-concentrat
756	5	2.3	119	2	JC4546	hypothetical prote	829	5	2.3	132	2	S34653	melanin-concentrat
757	5	2.3	120	1	E37854	dihydroneopterin a	830	5	2.3	132	2	S34654	melanin-concentrat
758	5	2.3	120	2	C29775	Ig kappa chain pre	831	5	2.3	132	2	A32910	melanin-concentrat
759	5	2.3	120	2	E29775	Ig kappa chain pre	832	5	2.3	132	2	S23374	T-cell receptor al

833	5	2.3	132	2	T17596	906	5	2.3	141	2	H71504	ribosomal protein
834	5	2.3	133	1	R5EG16	907	5	2.3	141	2	JQ1533	C2 protein - Panic
835	5	2.3	133	2	A23725	908	5	2.3	141	2	T38831	hypothetical prote
836	5	2.3	133	2	S78365	909	5	2.3	141	2	S29983	lectin II - Geodia
837	5	2.3	133	2	JA0087	910	5	2.3	142	2	C64071	H+-transporting AT
838	5	2.3	133	2	S57885	911	5	2.3	142	2	F71030	hypothetical prote
839	5	2.3	133	2	G45893	912	5	2.3	142	2	S39708	ywC protein - Bac
840	5	2.3	133	2	S6285	913	5	2.3	143	1	EHUND	Ig heavy chain pre
841	5	2.3	133	2	A41298	914	5	2.3	143	1	R3R211	ribosomal protein
842	5	2.3	133	2	S47700	915	5	2.3	143	2	D71911	ribonuclease hi -
843	5	2.3	133	2	T14238	916	5	2.3	143	2	E64602	ribonuclease H - H
844	5	2.3	133	2	PH0210	917	5	2.3	143	2	S58584	ribosomal protein
845	5	2.3	133	2	E59069	918	5	2.3	143	2	S16248	heat shock protein
846	5	2.3	133	2	I84500	919	5	2.3	143	2	A70760	hypothetical prote
847	5	2.3	133	2	F71189	920	5	2.3	144	2	T07175	ferredoxin [2Fe-2S
848	5	2.3	134	2	T07357	921	5	2.3	144	2	A29101	vasopressin / neur
849	5	2.3	134	2	T34591	922	5	2.3	144	2	E64156	hypothetical prote
850	5	2.3	135	2	T03288	923	5	2.3	144	2	G71959	neutrophil-activat
851	5	2.3	135	2	JH0684	924	5	2.3	144	2	C64550	neutrophil activat
852	5	2.3	135	2	JH0188	925	5	2.3	144	2	A70500	hypothetical prote
853	5	2.3	135	2	S22591	926	5	2.3	144	2	C69527	methylmalonyl-CoA
854	5	2.3	135	2	S39212	927	5	2.3	144	2	T33109	hypothetical prote
855	5	2.3	135	2	G70519	928	5	2.3	145	2	S40831	hypothetical 15.9K
856	5	2.3	135	2	A70317	929	5	2.3	145	2	S77196	hypothetical prote
857	5	2.3	135	2	F72688	930	5	2.3	145	2	C49852	host-lethality pro
858	5	2.3	136	2	B21105	931	5	2.3	145	2	G75485	conserved hypothet
859	5	2.3	136	2	S22977	932	5	2.3	145	2	T21691	hypothetical prote
860	5	2.3	136	2	S31695	933	5	2.3	145	2	T33961	hypothetical prote
861	5	2.3	136	2	G72670	934	5	2.3	145	2	C71378	hypothetical prote
862	5	2.3	136	2	A70142	935	5	2.3	146	2	B25929	hemoglobin beta ch
863	5	2.3	136	2	S78390	936	5	2.3	147	1	ICRT4	interleukin-4 prec
864	5	2.3	136	2	S46790	937	5	2.3	147	1	QVZCP	F7 protein - sheep
865	5	2.3	136	2	S59638	938	5	2.3	147	2	J50148	obesity protein -
866	5	2.3	136	2	T10795	939	5	2.3	147	2	A70462	ribosomal protein
867	5	2.3	136	2	E71883	940	5	2.3	147	2	JQ1134	hypothetical prote
868	5	2.3	136	2	G75281	941	5	2.3	147	2	H75319	hypothetical prote
869	5	2.3	136	2	T22797	942	5	2.3	147	2	S15487	methylmalonyl-CoA
870	5	2.3	137	1	IGP1	943	5	2.3	147	2	A59866	hypothetical prote
871	5	2.3	137	1	AVSP	944	5	2.3	147	2	A59198	hypothetical prote
872	5	2.3	137	2	S22515	945	5	2.3	147	2	T04061	hypothetical prote
873	5	2.3	137	2	S07648	946	5	2.3	147	2	T09621	membrane channel p
874	5	2.3	137	2	B72335	947	5	2.3	147	2	T20615	hypothetical prote
875	5	2.3	137	2	S38478	948	5	2.3	147	2	T29241	hypothetical prote
876	5	2.3	137	2	D38163	949	5	2.3	147	2	A4580	hypothetical prote
877	5	2.3	137	2	T07694	950	5	2.3	148	2	S40148	gastrula-specific
878	5	2.3	137	2	G70346	951	5	2.3	148	2	S39403	integrin alpha-7A
879	5	2.3	138	1	R3NT11	952	5	2.3	148	2	E69151	hupC protein - Bra
880	5	2.3	138	1	R3PM11	953	5	2.3	148	2	T39949	formate hydrogenly
881	5	2.3	138	1	R3HU2	954	5	2.3	149	2	T04171	probable copper tr
882	5	2.3	138	1	Q8EB2	955	5	2.3	149	2	A71074	heat shock protein
883	5	2.3	138	2	S59086	956	5	2.3	149	2	S12192	probable methymal
884	5	2.3	138	2	A42495	957	5	2.3	149	2	D69401	hypothetical prote
885	5	2.3	138	2	B27863	958	5	2.3	149	2	S13551	hypothetical prote
886	5	2.3	138	2	F75120	959	5	2.3	150	2	T03286	stem-specific prot
887	5	2.3	138	2	S67240	960	5	2.3	150	2	S20874	ferredoxin [2Fe-2S
888	5	2.3	138	2	S64058	961	5	2.3	150	2	S20874	heat shock protein
889	5	2.3	138	2	G69191	962	5	2.3	150	2	F71473	heat shock protein
890	5	2.3	139	1	Q2BP87	963	5	2.3	150	2	B75296	probable hydrolase
891	5	2.3	139	2	A41287	964	5	2.3	150	2	D75354	hypothetical prote
892	5	2.3	139	2	S31443	965	5	2.3	150	2	F75070	hypothetical prote
893	5	2.3	139	2	F73228	966	5	2.3	151	1	S36423	ribosomal protein
894	5	2.3	139	2	I53298	967	5	2.3	151	1	A84359	ribosomal protein
895	5	2.3	139	2	T18449	968	5	2.3	151	1	HHWT17	heat shock protein
896	5	2.3	139	2	C71731	969	5	2.3	151	2	S21600	heat shock protein
897	5	2.3	139	2	T17378	970	5	2.3	151	2	T05739	probable heat shoc
898	5	2.3	139	2	B75307	971	5	2.3	151	2	E72806	gp55 protein - Myc
899	5	2.3	139	2	F70892	972	5	2.3	151	2	T14714	reverse transcript
900	5	2.3	139	2	T41526	973	5	2.3	151	2	D70007	hypothetical prote
901	5	2.3	140	1	WMVZF4	974	5	2.3	151	2	G70903	hypothetical prote
902	5	2.3	140	2	T01170	975	5	2.3	151	2	B70862	ferredoxin [2Fe-2S
903	5	2.3	140	2	C64651	976	5	2.3	151	2	T38838	hypothetical prote
904	5	2.3	140	2	T07753	977	5	2.3	151	2	S37812	hypothetical prote
905	5	2.3	140	2	T14771	978	5	2.3	152	1	AWRT	atrial natriuretic

A; Molecule type: DNA

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033

A:Accession: A69943

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-229 <KUN>

A:Cross-references: GB:299116; GB:AL009126; NID:92634723; PIDN:CAB14248.1; PID:el185585;

A:Experimental source: strain 168

C:Genetics:

A:Gene: ypuL

C:Superfamily: conserved hypothetical protein HI1243

Query Match 4.5%; Score 10; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 0.008;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 EGRNQVRM 185

Db 184 EGRNQVRM 193

RESULT 5

D75462

ribosomal large subunit pseudouridine synthase B - *Deinococcus radiodurans* (strain R1)

C:Species: *Deinococcus radiodurans*

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: D75462

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250

A:Accession: D75462

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-257 <WHI>

A:Cross-references: GB:AE001942; GB:AE000513; NID:96458611; PIDN:AAF10472.1; PID:9645861

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0896

A:Map position: 1

Query Match 4.5%; Score 10; DB 2; Length 257;

Best Local Similarity 100.0%; Pred. No. 0.0089;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 EGRNQVRM 185

Db 181 EGRNQVRM 190

RESULT 6

T19185

hypothetical protein C10C6.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19185

R:White, S.

submitted to the EMBL Data Library, December 1996

A:Reference number: Z19086

A:Accession: T19185

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-341 <WIL>

A:Cross-references: EMBL:283217; PIDN:CAB05680.1; GSPDB:GN00022; CESP:C10C6.3

A:Experimental source: clone C10C6

C:Genetics:

A:Gene: CESP:C10C6.3

A:Map position: 4

A:Introns: 14/3; 81/2; 137/2; 211/2; 261/3

Query Match 3.6%; Score 8; DB 2; Length 341;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 VOVEGEPD 120

Db 45 VOVEGEPD 52

RESULT 7

A1347

hypothetical protein TP0260 - *syphilis spirochete*

C:Species: *Treponema pallidum* subsp. *pallidum* (*syphilis spirochete*)

C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 24-Nov-1999

C:Accession: A71347

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of *Treponema pallidum*, the *syphilis spirochete*.

A:Reference number: A71250; MUID:98332770

A:Accession: A71347

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-448 <CO>

A:Cross-references: GB:AE001207; GB:AE000520; NID:93322526; PIDN:AAC65253.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0260

C:Superfamily: *syphilis spirochete* hypothetical protein TP0260

Query Match 3.6%; Score 8; DB 2; Length 448;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 QFTDEAGR 62

Db 287 QFTDEAGR 294

RESULT 8

B41492

58K antigen - *Rickettsia tsutsugamushi*

C:Species: *Rickettsia tsutsugamushi*

C>Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 26-Aug-1999

C:Accession: B41492

R:Stover, C.K.; Marana, D.P.; Dasch, G.A.; Oaks, E.V.

Infect. Immun. 58, 1360-1368, 1990

A:Title: Molecular cloning and sequence analysis of the *Sta58* major antigen gene of *R*

proteins.

A:Reference number: A41492; MUID:90216005

A:Accession: B41492

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-555 <STO>

A:Cross-references: GB:M1887; NID:9152501; PIDN:AAA26393.1; PID:9152503

C:Superfamily: *chaperonin groEL*

Query Match 3.6%; Score 8; DB 2; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 LAKLRNGV 131

Db 372 LAKLRNGV 379

RESULT 9

B75621
sensor histidine kinase/response regulator - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: B75621
A:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Otterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250
A:Accession: B75621
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <WHI>
A:Cross-references: GB:AF001826; NID:g6460827; PIDN:AAF12582.1; PID:g6460878; TIGR:DRB00
C:Genetics:
A:Map position: megaplasmid
A:Gene: DRB0028
A:Genome: plasmid
A:Note: plasmid MP1

Query Match 3.6%; Score 8; DB 2; Length 577;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 DEAGRSTL 65
|||||
Db 64 DEAGRSTL 71

RESULT 10

JE0169
dnaJ heat shock protein MCG18 - human
A:Alternate names: HSPF2
C:Species: Homo sapiens (man)
C:Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C:Accession: JE0169
R:Sillins, G.; Grimmond, S.; Hayward, N.
Biochem. Biophys. Res. Commun. 243, 273-276, 1998
A:Title: Characterisation of a new human and murine member of the DnaJ family of proteins
A:Reference number: JE0169; MUID:98139904
A:Accession: JE0169
A:Molecule type: mRNA
A:Residues: 1-241 <STL>
A:Cross-references: GB:AF036874
C:Genetics:

A:Gene: MCG18
A:Map position: 11q13
A:Introns: 29/2 60/3 176/2 205/2
C:Superfamily: dnaJ amino-terminal homology
F:34-102/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 3.2%; Score 7; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 152 LWPRNPP 158
|||||
Db 12 LWPRNPP 18

RESULT 11

A43662
replicating protein repA - Agrobacterium tumefaciens plasmid pTAR
C:Species: Agrobacterium tumefaciens
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A43662
R:Gallie, D.R.; Kado, C.I.
J. Bacteriol. 170, 3170-3176, 1988

A:Title: Minimal region necessary for autonomous replication of pTAR.
A:Reference number: A43662; MUID:88257036
A:Accession: A43662
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-250 <GAL>
A:Cross-references: GB:M21299; NID:g142263; PIDN:AAI5307.1; PID:g142266
C:Genetics:
A:Genome: plasmid

Query Match 3.2%; Score 7; DB 2; Length 250;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 QPKRRTG 108
|||||
Db 59 QPKRRTG 65

RESULT 12

S41382
pobR protein - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: S41382
R:Entsch, B.; Squire, L.; Wicks, R.E.
submitted to the EMBL Data Library, December 1993
A:Description: Gene for the regulation of PARA-hydroxybenzoate hydroxylase in Pseudom
A:Reference number: S41380
A:Accession: S41382
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-288 <ENT>
A:Cross-references: EMBL:X76994; NID:g444024; PIDN:CAA54301.1; PID:g444027

Query Match 3.2%; Score 7; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 VQVEGEP 119
|||||
Db 63 VQVEGEP 69

RESULT 13

S74669
hypothetical protein sl11664 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S74669
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S74669
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <KAN>
A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651769; PIDN:BAAL6821.1; PID:d101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 3.2%; Score 7; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 71 VQGVYAA 77
|||||

Db 298 VQGYAA 304

RESULT 14

A71965
hypothetical protein jhp0165 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: A71965
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: A71965
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-406 <ARN>
A:Cross-references: GB:AE001455; GB:AE001439; NID:g4154678; PIDN:AAD05746.1; PID:g415468
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0165

Query Match 3.2%; Score 7; DB 2; Length 406;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 ANGEWRD 216

|||||

Db 356 ANGEWRD 362

RESULT 15

T35794
Probable 4-aminobutyrate aminotransferase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 03-Dec-1999
C:Accession: T35794
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21556
A:Accession: T35794
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-444 <MUR>
A:Cross-references: EMBL:AL031225; PIDN:CAA20213.1; GSPDB:GN00070; SCOEDB:SC8B7.02
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: gabt; SCOEDB:SC8B7.02

Query Match 3.2%; Score 7; DB 2; Length 444;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 EGLLVLT 91

|||||

Db 403 EGLLVLT 409

Search completed: June 10, 2000, 11:27:08
Job time: 2521 sec

THIS PAGE BLANK (REF ID: A67700)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 11:27:09 ; Search time 9.96 Seconds
(without alignments)
678.816 Million cell updates/sec

Title: US-09-252-691-7056

Perfect score: 222

Sequence: 1 AIMQLTPTWNTKTSFRK.....SYTLDLSENGEWRDVTPEKN 222

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 83857 seqs, 30454973 residues

Word size : 0

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	23.4	217	1 YMF_CECOLI	P75966 escherichia
2	15	6.8	240	1 YMF_CECOLI	P44827 haemophilus
3	12	5.4	261	1 Y612_SYNY3	P72581 synechocyst
4	10	4.5	229	1 RLUB_BACSU	P35159 bacillus su
5	8	3.6	555	1 CH60_RICTS	P16525 rickettsia
6	7	3.2	250	1 REPA_AGRU	P15394 agrobacteri
7	7	3.2	387	1 STCV_EMENI	Q00727 emericeila
8	7	3.2	454	1 NEM_PIG	P08552 sus scrofa
9	7	3.2	515	1 PDI_ASPNG	Q12730 aspergillus
10	7	3.2	534	1 TCP2_YEAST	P39077 saccharomyc
11	7	3.2	546	1 TCP2_YEAST	P39079 saccharomyc
12	7	3.2	604	1 SP20_YEAST	P50875 saccharomyc
13	7	3.2	707	1 COG9_RABIT	P41246 oryctolagus
14	7	3.2	764	1 ACOC_CUCMC	Q42659 cucumis mel
15	7	3.2	845	1 NEM_RAT	P12839 rattus norv
16	7	3.2	848	1 NEM_MOUSE	P08553 mus musculu
17	7	3.2	898	1 ACOC_CUCMA	P49608 cucurbita m
18	7	3.2	915	1 NFM_HUMAN	P07197 homo sapien
19	7	3.2	990	1 TRAT_ECOLI	P13694 escherichia
20	7	3.2	1148	1 MFD_ECOLI	P30958 escherichia
21	7	3.2	1886	1 GP21_RAT	P11654 rattus norv
22	6	2.7	89	1 VFUS_ORFN2	P26654 orf virus (
23	6	2.7	94	1 IHFB_PSEAE	Q51473 pseudomonas
24	6	2.7	97	1 Y109_EMENI	P05678 emericeila
25	6	2.7	103	1 CEAV_ECOLI	P22522 escherichia
26	6	2.7	106	1 YNTU_RHOCA	Q07184 rhodobacter
27	6	2.7	133	1 IL4_PIG	Q04745 sus scrofa
28	6	2.7	134	1 HI_EUPEU	P32103 euplotes eu
29	6	2.7	142	1 YFCC_CHRV1	Q06536 chromatiu
30	6	2.7	150	1 FKBX_PSEFL	P21863 pseudomonas
31	6	2.7	158	1 UREE_KLEAE	P18317 klebsiella
32	6	2.7	165	1 LKTC_PASSP	P55124 pasteurella
33	6	2.7	173	1 CRGD_HUMAN	P07320 homo sapien

P42570 drosophila
P07096 petromyzon
P29091 schizophyll
P44869 haemophilus
P02522 bos taurus
O35943 mus musculu
P45270 haemophilus
Q16395 homo sapien
P09629 homo sapien
P09024 mus musculu
P33918 escherichia
Q53117 rhodococcus
P05324 methanococc
P50908 thermocoga
P05325 methanococc
P70973 bacillus su
Q55578 synechocyst
P54450 bacillus su
P07283 saccharomyc
P09551 escherichia
Q58515 methanococc
P50363 allomyces a
P50364 allomyces m
P38440 ovis aries
P35895 escherichia
Q06778 pseudomonas
P14489 pseudomonas
Q07400 mycobacteri
P08164 bacillus su
P71671 mycobacteri
P83375 saccharomyc
P53964 saccharomyc
P43468 pediococcus
Q07203 mycobacteri
Q59281 corynebacte
Q44104 amycolatops
P02723 neurospora
P37415 salmonella
Q58350 methanococc
P31463 escherichia
P17264 rana catesb
P02532 rana tempor
P14061 homo sapien
Q09704 schizosacch
P38749 saccharomyc
P52559 brucella me
Q11030 mycobacteri
P14187 neurospora
P16594 rice dwarf
P24900 salmonella
P03016 escherichia
Q08142 podospora a
P43767 haemophilus
P34258 caenorhabdi
P07994 bos taurus
Q53411 mycobacteri
P22839 proteus mir
P46545 lactobacill
Q40353 medicago sa
Q11037 mycobacteri
Q14791 homo sapien
Q58431 methanococc
P41052 schizosacch
Q40902 schizosacch
P10899 mycobacteri
P22044 human parai
Q01889 trypanosoma
P72052 mycobacteri
Q05514 clostridium
P24531 campylobact
Q01604 drosophila
Q57017 haemophilus
Q49900 mycobacteri

107	6	2.7	428	1	CG2B_ARATH	P30183 arabidopsis	180	6	2.7	669	1	GSH1_SCHPO	O09768 schizosacch
108	6	2.7	432	1	FIBX_MOUSE	P12804 mus musculus	181	6	2.7	677	1	SG1_MOUSE	P16014 mus musculus
109	6	2.7	450	1	YABL_ECOLI	P37764 escherichia	182	6	2.7	677	1	YPK2_YEAST	P18961 saccharomyc
110	6	2.7	451	1	ARGD_ALNGL	O04866 alnus glut	183	6	2.7	709	1	SYR_CAEEL	Q19825 caenorhabdi
111	6	2.7	462	1	YP59_MYCTU	O50739 mycobacteri	184	6	2.7	722	1	MESD_LEUME	Q10418 leuconostoc
112	6	2.7	465	1	DCDR_XENIA	P42291 xenopus lae	185	6	2.7	746	1	GUNI_STRRE	O05156 streptomyc
113	6	2.7	466	1	BIAR_MOUSE	P34971 mus musculus	186	6	2.7	775	1	VP4_ROT3	P39033 human rotav
114	6	2.7	466	1	BIAR_MOUSE	P34971 mus musculus	187	6	2.7	775	1	VP4_ROT3	P39033 human rotav
115	6	2.7	467	1	PPAR_CAVPO	P18090 rattus norv	188	6	2.7	779	1	GEF1_YEAST	P370214 human rotav
116	6	2.7	474	1	DLD2_BACSU	O35507 cavia porce	189	6	2.7	810	1	MECB_BACSU	P37571 bacillus su
117	6	2.7	488	1	DHAL_BACST	P45333 bacillus st	190	6	2.7	823	1	YHX2_YEAST	P38863 saccharomyc
118	6	2.7	499	1	VLI_HPV13	O02273 human papil	191	6	2.7	847	1	HEX_VIBVO	O04786 vibrio vuln
119	6	2.7	499	1	VLI_HPV33	O06415 human papil	192	6	2.7	856	1	PPSA_AQUAE	O67899 aquifex aeo
120	6	2.7	499	1	VLI_HPV53	O05113 human papil	193	6	2.7	859	1	LUXQ_VIBHA	P54302 vibrio harv
121	6	2.7	500	1	CPJ1_RABIT	P52786 oryctolagus	194	6	2.7	862	1	LOX1_HORVU	P29114 hordeum vul
122	6	2.7	500	1	VLI_HPV6A	P03100 human papil	195	6	2.7	863	1	APCE_CVACA	P35511 cyanidium c
123	6	2.7	501	1	ATPA_STRMU	P95787 streptococ	196	6	2.7	865	1	LOX2_ORISA	P29250 oryza sativ
124	6	2.7	501	1	VLI_HPV11	P04012 human papil	197	6	2.7	869	1	PEM1_YEAST	P05374 saccharomyc
125	6	2.7	501	1	VLI_HPV1	P22163 rhesus papil	198	6	2.7	878	1	IL3B_MOUSE	P26954 mus musculus
126	6	2.7	502	1	VLI_HPV28	P30791 human papil	199	6	2.7	879	1	PPRP_RAT	O62786 rattus norv
127	6	2.7	502	1	VLI_HPV42	P27233 human papil	200	6	2.7	896	1	CYRB_MOUSE	P26955 mus musculus
128	6	2.7	502	1	VLI_FCPV1	O02274 pygmy chimp	201	6	2.7	897	1	CYRB_MOUSE	P26955 mus musculus
129	6	2.7	503	1	VLI_HPV26	P36735 human papil	202	6	2.7	909	1	YBPE_SCHPO	P32327 homo sapien
130	6	2.7	503	1	VLI_HPV66	O08061 human papil	203	6	2.7	910	1	HUL5_YEAST	O09879 schizosacch
131	6	2.7	504	1	VLI_HPV70	P50793 human papil	204	6	2.7	914	1	YK69_YEAST	P36165 saccharomyc
132	6	2.7	505	1	VLI_HPV39	P48838 human papil	205	6	2.7	923	1	GUX2_CLOS	P50900 clostridium
133	6	2.7	505	1	VLI_HPV61	P50822 human papil	206	6	2.7	923	1	AGLU_TETPY	O00906 tetrahymena
134	6	2.7	505	1	VLI_HPV68	P54669 human papil	207	6	2.7	946	1	RGCI_HUMAN	P98171 homo sapien
135	6	2.7	505	1	VLI_HPVME	P27964 human papil	208	6	2.7	947	1	PODK_MAIZE	P11155 zea mays (m
136	6	2.7	508	1	VLI_HPV30	O02515 human papil	209	6	2.7	949	1	PODK_MESCR	O42910 mesembryant
137	6	2.7	510	1	VLI_HPV2A	P25486 human papil	210	6	2.7	953	1	PODK_FLABI	Q39735 flaveria bi
138	6	2.7	510	1	VLI_HPV57	P22162 human papil	211	6	2.7	953	1	PODK_FLATR	P22221 flaveria tr
139	6	2.7	515	1	CALR_MOUSE	P26679 enterococc	212	6	2.7	955	1	PODK_FLABR	Q39734 flaveria br
140	6	2.7	515	1	CATA_YEAST	P15202 saccharomyc	213	6	2.7	956	1	PODK_FLAPR	Q42736 flaveria pr
141	6	2.7	518	1	ATPA_ENTHR	P26679 enterococc	214	6	2.7	971	1	VP2_EHDV1	O00273 epizootic h
142	6	2.7	522	1	CYSJ_THIRO	P52674 thiocapsa r	215	6	2.7	979	1	VGLB_HSVEL	P25218 equine herp
143	6	2.7	524	1	VLI_HPV58	P26535 human papil	216	6	2.7	980	1	VGLB_HSVE1	P18050 equine herp
144	6	2.7	529	1	VLI_HPV52	O05138 human papil	217	6	2.7	980	1	VGLB_HSVE1	P18050 equine herp
145	6	2.7	531	1	VLI_HPV03	P36732 human papil	218	6	2.7	980	1	VGLB_HSVE1	P18050 equine herp
146	6	2.7	532	1	VLI_HPV10	P36731 human papil	219	6	2.7	1008	1	GI10_YEAST	P28922 equine herp
147	6	2.7	534	1	VLI_HPV56	P36743 human papil	220	6	2.7	1033	1	YDK9_SCHPO	P36048 saccharomyc
148	6	2.7	539	1	VLI_HPV45	P36741 human papil	221	6	2.7	1039	1	SYL_METJA	P87115 schizosacch
149	6	2.7	544	1	4CL2_PETCR	P14913 desulfoselinu	222	6	2.7	1081	1	KOKO_YEAST	O58357 methanococc
150	6	2.7	544	1	PRIS_DESDE	O01770 desulfovibr	223	6	2.7	1123	1	PHYA_SOLTU	Q12236 saccharomyc
151	6	2.7	547	1	ILVD_ARCFU	O29248 archaeoglob	224	6	2.7	1124	1	PHY1_TOBAC	P30733 solanum tub
152	6	2.7	548	1	ATPA_KLULA	P49375 kluyveromyc	225	6	2.7	1144	1	RIR1_HSV23	P33530 nicotiana t
153	6	2.7	548	1	PRIS_METJA	O58175 methanococc	226	6	2.7	1157	1	Y182_HUMAN	P09853 herpes simp
154	6	2.7	553	1	PRIS_DESVH	P31101 desulfovibr	227	6	2.7	1174	1	KCRF_STRPU	Q14687 homo sapien
155	6	2.7	553	1	SYQ_ECOLI	P00962 escherichia	228	6	2.7	1179	1	RPOB_MYCDE	P18294 strongyloce
156	6	2.7	554	1	UL25_HSV7J	P52388 herpes simp	229	6	2.7	1227	1	LAF4_HUMAN	P30760 mycobacteri
157	6	2.7	554	1	YDA9_SCHPO	Q10351 schizosacch	230	6	2.7	1239	1	CHS5_USTMA	P51826 homo sapien
158	6	2.7	554	1	YG62_ARATH	O22212 arabidopsis	231	6	2.7	1242	1	IRS1_HUMAN	O13394 ustilago ma
159	6	2.7	556	1	PRIS_THIFE	P96095 thiobacillu	232	6	2.7	1290	1	PIP4_HUMAN	P35568 homo sapien
160	6	2.7	559	1	TCPG_OXYGR	O00782 oxytricha q	233	6	2.7	1290	1	PIP4_RAT	P19174 homo sapien
161	6	2.7	568	1	VLI_HPV18	P06794 human papil	234	6	2.7	1291	1	PIP4_BOVIN	P10686 rattus norv
162	6	2.7	573	1	CATT_YEAST	O06115 saccharomyc	235	6	2.7	1330	1	VCAP_PPRIS	P08487 bos taurus
163	6	2.7	575	1	RECJ_ERWCH	P39693 erwania chr	236	6	2.7	1344	1	IF3A_MOUSE	Q00705 pseudorabie
164	6	2.7	575	1	RECJ_HAEIN	P45112 haemophilus	237	6	2.7	1382	1	IF3A_HUMAN	P23116 mus musculus
165	6	2.7	577	1	RECJ_ECOLI	P21893 escherichia	238	6	2.7	1394	1	CNG4_BOVIN	O14152 homo sapien
166	6	2.7	584	1	VLI_HPV27	P36736 human papil	239	6	2.7	1475	1	TRA2_CAEEL	Q28181 bos taurus
167	6	2.7	600	1	LEPA_RICPR	O92861 rickettsia	240	6	2.7	1481	1	APU_THEET	P38939 t amyllopull
168	6	2.7	602	1	PSY_NEUCR	P37395 neurospora	241	6	2.7	1497	1	TRA2_CAEER	Q17307 caenorhabdi
169	6	2.7	606	1	PF20_CHIRE	P93107 chlamydomon	242	6	2.7	1515	1	GDE_HUMAN	P35573 homo sapien
170	6	2.7	613	1	FIBP_ADEMI	P19721 mouse adeno	243	6	2.7	1555	1	GDE_RABIT	P35574 oryctolagus
171	6	2.7	614	1	NTBE_HUMAN	P48065 homo sapien	244	6	2.7	1567	1	YH81_YEAST	Q04781 saccharomyc
172	6	2.7	614	1	NTBE_HUMAN	P48065 homo sapien	245	6	2.7	1567	1	YH81_YEAST	Q04781 saccharomyc
173	6	2.7	626	1	PM17_MOUSE	O60696 mus musculus	246	6	2.7	1842	1	ICEN_XANCT	P18127 xanthomonas
174	6	2.7	646	1	FATP_MOUSE	Q60714 mus musculus	247	6	2.7	1859	1	RPB1_CAEEL	Q10289 schizosacch
175	6	2.7	655	1	CD44_MOUSE	P15379 mus musculus	248	6	2.7	2073	1	BIME_EMENI	P16356 caenorhabdi
176	6	2.7	657	1	PRIA_TREPA	O83358 treponema p	249	6	2.7	2205	1	POLG_POL2W	P24686 emericeella
177	6	2.7	657	1	YGG6_YEAST	P53165 saccharomyc	250	6	2.7	2206	1	POLG_POL2W	P23069 poliovirus
178	6	2.7	661	1	PM17_HUMAN	P40967 homo sapien	251	6	2.7	2206	1	POLG_POL2W	P23055 coxsackiev1
179	6	2.7	665	1	NUC2_SCHPO	P10505 schizosacch	252	6	2.7	2206	1	POLG_POL2W	P06209 poliovirus

180	6	2.7	669	1	GSH1_SCHPO	O09768 schizosacch
181	6	2.7	677	1	SG1_MOUSE	P16014 mus musculus
182	6	2.7	677	1	YPK2_YEAST	P18961 saccharomyc
183	6	2.7	709	1	SYR_CAEEL	Q19825 caenorhabdi
184	6	2.7	722	1	MESD_LEUME	Q10418 leuconostoc
185	6	2.7	746	1	GUNI_STRPRE	O05156 streptomyc
186	6	2.7	775	1	VP4_ROT3	P39033 human rotav
187	6	2.7	775	1	VP4_ROT3	P39033 human rotav
188	6	2.7	779	1	GEF1_YEAST	P370214 human rotav
189	6	2.7	810	1	MECB_BACSU	P37571 bacillus su
190	6	2.7	823	1	YHX2_YEAST	P38863 saccharomyc
191	6	2.7	847	1	HEX_VIBVO	O04786 vibrio vuln
192	6	2.7	856	1	PPSA_AQUAE	O67899 aquifex aeo
193	6	2.7	859	1	LUXQ_VIBHA	P54302 vibrio harv
194	6	2.7	862	1	LOX1_HORVU	P29114 hordeum vul
195	6	2.7	863	1	APCE_CVACA	P35911 cyanidium c
196	6	2.7	865	1	LOX2_ORISA	P29250 oryza sativ
197	6	2.7	869	1	PEM1_YEAST	P05374 saccharomyc
198	6	2.7	878	1	IL3B_MOUSE	P26954 mus musculus
199	6	2.7	879	1	PPRP_RAT	O62786 rattus norv
200	6	2.7	896	1	CYRB_MOUSE	P26955 mus musculus
201	6	2.7	897	1	CYRB_MOUSE	P26955 mus musculus
202	6	2.7	909	1	UBBP_SCHPO	P32927 homo sapien
203	6	2.7	910	1	HUL5_YEAST	O09879 schizosacch
204	6	2.7	914	1	YK69_YEAST	P36165 saccharomyc
205	6	2.7	914	1	GUX2_CLOS	P05900 clostridium
206	6	2.7	923	1	AGLU_TETPY	O00906 tetrahymena
207	6	2.7	946	1	RGCL_HUMAN	P18171 homo sapien
208	6	2.7	947	1	PODK_WAIZE	P11557 zea mays (m
209	6	2.7	949	1	PODK_MESCI	P42910 mesembryant
210	6	2.7	953	1	PODK_FLABI	Q39735 flaveria bi
211	6	2.7	953	1	PODK_FLATYR	P22221 flaveria tr
212	6	2.7	955	1	PODK_FLABR	Q39734 flaveria br
213	6	2.7	956	1	PODK_FLAPR	Q42736 flaveria pr
214	6	2.7	971	1	VP2_EDHVL	Q00273 epizootic h
215	6	2.7	979	1	VGLB_HSVEL	P25218 equine herp
216	6	2.7	980	1	VGLB_HSVBL	P18050 equine herp
217	6	2.7	980	1	VGLB_HSVFA	P18551 equine herp
218	6	2.7	980	1	VGLB_HSVFB	P28922 equine herp
219	6	2.7	1008	1	G110_YEAST	P36048 saccharomyc
220	6	2.7	1033	1	YDK9_SCHPO	P87115 schizosacch
221	6	2.7	1039	1	SYL_METUA	P08648 bos sapien
222	6	2.7	1081	1	KOKO_YEAST	P30760 mycobacteri
223	6	2.7	1123	1	PHYA_SOLTU	P18294 strongyloce
224	6	2.7	1124	1	PHY1_TOBAC	P30760 mycobacteri
225	6	2.7	1144	1	RRL_HSV23	Q12236 saccharomyc
226	6	2.7	1157	1	I182_HUMAN	P30733 solanum tub
227	6	2.7	1174	1	KGRF_STRPU	P33530 nicotiana t
228	6	2.7	1179	1	ROB_MYCLE	P09853 herpes simp
229	6	2.7	1227	1	LAF4_HUMAN	Q14687 homo sapien
230	6	2.7	1239	1	CHS5_USTMA	P18294 strongyloce
231	6	2.7	1242	1	IRS1_HUMAN	P30760 mycobacteri
232	6	2.7	1290	1	PIP4_HUMAN	P18294 strongyloce
233	6	2.7	1290	1	PIP4_HUMAN	P18294 strongyloce
234	6	2.7	1291	1	PIP4_RAT	P18294 strongyloce
235	6	2.7	1330	1	PIP4_BOVIN	P18294 strongyloce
236	6	2.7	1344	1	ICAP_PRLVIS	P30760 mycobacteri
237	6	2.7	1382	1	IF3A_MOUSE	P18294 strongyloce
238	6	2.7	1394	1	IF3A_HUMAN	P18294 strongyloce
239	6	2.7	1475	1	TRA2_CAEEL	P18294 strongyloce
240	6	2.7	1481	1	APU_THEET	P18294 strongyloce
241	6	2.7	1497	1	TRA2_CAEER	P18294 strongyloce
242	6	2.7	1515	1	GDE_HUMAN	P18294 strongyloce
243	6	2.7	1555	1	GDE_RABIT	P18294 strongyloce
244	6	2.7	1562	1	YH81_YEAST	P18294 strongyloce
245	6	2.7	1567	1	ICEN_XANTC	P18294 strongyloce
246	6	2.7	1842	1	FAS2_SCHPO	P18294 strongyloce
247	6	2.7	1859	1	RBP1_CAEEL	P18294 strongyloce
248	6	2.7	2073	1	BIME_EMENT	P18294 strongyloce
249	6	2.7	2205	1	POLG_POL2W	P18294 strongyloce
250	6	2.7	2206	1	POLG_CXA21	P18294 strongyloce
251	6	2.7	2206	1	POLG_POL12	P18294 strongyloce
252	6	2.7	2206	1	POLG_POL32	P18294 strongyloce

253	6	2.7	2206	1	POLG_POL3L	P03302 poliovirus	326	5	2.3	97	1	EOTA_MOUSE	P48298 mus musculu
254	6	2.7	2207	1	POLG_POL2L	P06210 poliovirus	327	5	2.3	97	1	EOTA_RAT	P97545 rattus norv
255	6	2.7	2208	1	POLH_POL1M	P03300 p genome po	328	5	2.3	97	1	KFER_ARCLA	P00223 artium lap
256	6	2.7	2209	1	POLG_POL1S	P03301 poliovirus	329	5	2.3	97	1	KFER_CHICK	P20308 gallus gall
257	6	2.7	2214	1	POLG_CXA24	P36290 coxsackievi	330	5	2.3	97	1	MAII_SHIFL	Q06080 shigella fl
258	6	2.7	2261	1	RRLP_HUMPM	P30929 mumps virus	331	5	2.3	99	1	Y11K_STRFR	P26800 streptomyc
259	6	2.7	2670	1	YAO5_SCHPO	Q10105 schizosacch	332	5	2.3	99	1	Y14L_METJA	Q57606 methanococc
260	6	2.7	3011	1	POLG_HCVH	P27958 h genome po	333	5	2.3	100	1	2SS4_CAPMA	P80353 capparis ma
261	6	2.7	3092	1	IRAL1_YEAST	P18963 saccharomyc	334	5	2.3	100	1	REV_HV2CA	P24104 human immun
262	6	2.7	3110	1	HD_HUMAN	P51111 rattus norv	335	5	2.3	100	1	REV_HV2RO	P04615 human immun
263	6	2.7	3144	1	POLG_TBVEH	P42858 homo sapien	336	5	2.3	101	1	VNST_BUNYW	P16494 bnyamwera
264	6	2.7	3414	1	POLG_TBVEH	Q01299 t genome po	337	5	2.3	101	1	VNST_MAGV	P16605 maguari vir
265	6	2.7	3567	1	ERL2_SACER	Q03132 saccharopol	338	5	2.3	102	1	BL_BBV	P29099 black beetl
266	6	2.7	4543	1	LRP1_CHICK	P98157 gallus gall	339	5	2.3	102	1	G44A_DICDI	P16642 dictyostell
267	5	2.3	36	1	H115_ENSMI	P27203 ensis minor	340	5	2.3	102	1	ISS_ECOLI	P19592 escherichia
268	5	2.3	42	1	GBG7_MOUSE	Q61016 mus musculu	341	5	2.3	102	1	P1GR_PIG	Q29244 sus scrofa
269	5	2.3	43	1	BAGE_HUMAN	Q13072 homo sapien	342	5	2.3	102	1	RECA_ENTFA	P42444 enterococcu
270	5	2.3	48	1	DPSD_AZOVI	Q44558 azotobacter	343	5	2.3	102	1	RI10_GUITH	P19460 guillardia
271	5	2.3	48	1	THPO_HUMAN	P08918 homo sapien	344	5	2.3	102	1	TCPD_SCHPO	P50999 schizosacch
272	5	2.3	48	1	THPS_HUMAN	P08919 homo sapien	345	5	2.3	102	1	UN31_SCHPO	Q12238 schizosacch
273	5	2.3	49	1	THP1_BOVIN	P01249 bos taurus	346	5	2.3	102	1	YNI2_METTL	P03410 methanococc
274	5	2.3	49	1	THP2_BOVIN	P01250 bos taurus	347	5	2.3	103	1	G24B_DICDI	P16643 dictyostell
275	5	2.3	49	1	THPS_BOVIN	P01251 bos taurus	348	5	2.3	103	1	REV_HV2D1	P17754 human immun
276	5	2.3	49	1	Y195_RPT3	P10306 bacterioph	349	5	2.3	103	1	REV_HV2G1	P18039 human immun
277	5	2.3	53	1	RUBR_CHLLT	P09947 chlorobium	350	5	2.3	104	1	2SS3_CAPMA	P80352 capparis ma
278	5	2.3	55	1	RL15_LACLA	P27145 lactococcus	351	5	2.3	104	1	HKL5_ARATH	P48002 arabidopsis
279	5	2.3	60	1	HBX4_ECHGR	P55813 echinococcu	352	5	2.3	104	1	RECA_CLOAB	Q02441 clostridium
280	5	2.3	61	1	SECE_THETH	P38383 thermus aqu	353	5	2.3	104	1	RECA_LACDE	P02347 lactobacill
281	5	2.3	61	1	DNBI_BFDV	P13893 budgerigar	354	5	2.3	104	1	RECA_LACHE	Q02348 lactobacill
282	5	2.3	63	1	YM26_MARPO	P38468 marchantia	355	5	2.3	104	1	RECA_LEUME	Q02349 leuconostoc
283	5	2.3	68	1	GBG7_BOVIN	P30671 bos taurus	356	5	2.3	104	1	RECA_SCHPO	Q93374 schizosacch
284	5	2.3	68	1	GBG7_HUMAN	Q06262 homo sapien	357	5	2.3	104	1	URE2_MYCTU	P50048 mycobacteri
285	5	2.3	69	1	DRAG_PIG	Q29228 sus scrofa	358	5	2.3	106	1	GAC_YEAST	P33440 saccharomyc
286	5	2.3	69	1	GBG7_RAT	P43425 rattus norv	359	5	2.3	107	1	REV_HV2NZ	P29734 mastigoclad
287	5	2.3	69	1	Y375_TREPA	Q83390 treponema p	360	5	2.3	107	1	REV_HV2ST	P20870 human immun
288	5	2.3	70	1	PPG_RHIME	Q59752 rhizobium m	361	5	2.3	107	1	Y029_MYCPN	P75082 mycoplasma
289	5	2.3	70	1	RK31_PORPU	P51290 porphyra pu	362	5	2.3	108	1	YC50_CYAPA	P48326 cyanophora
290	5	2.3	70	1	RL31_HAEDU	Q59450 haemophilus	363	5	2.3	108	1	YNFA_ECOLI	P76169 escherichia
291	5	2.3	70	1	RL31_HAEIN	P44367 haemophilus	364	5	2.3	108	1	JHEA_TRINI	P30809 trichoplusi
292	5	2.3	70	1	Y4JN_RHISN	P55513 rhizobium s	365	5	2.3	109	1	VNST_BUNGE	P16992 bunyavirus
293	5	2.3	72	1	RK31_GUITH	Q46917 guillardia s	366	5	2.3	110	1	UCR6_BOVIN	P00129 bos taurus
294	5	2.3	72	1	RK31_ODOSI	P49563 odontella s	367	5	2.3	110	1	UCR6_HUMAN	P14927 homo sapien
295	5	2.3	73	1	RL29_SNY3	P73312 synecocyst	368	5	2.3	110	1	YD90_MYCTU	P71560 mycobacteri
296	5	2.3	74	1	HS12_SOYBN	P02520 glycine max	369	5	2.3	112	1	GLNB_MYCTU	Q10960 mycobacteri
297	5	2.3	76	1	RS18_NEIGO	Q07815 neisseria g	370	5	2.3	112	1	RAMA_ENTCL	P55922 enterobacte
298	5	2.3	76	1	YQFE_ECOLI	P52038 escherichia	371	5	2.3	113	1	RAMA_KLEPN	Q48413 klebsiella
299	5	2.3	77	1	VG43_BPM5	Q05255 mycobacteri	372	5	2.3	113	1	TRPF_BACME	P70938 bacillus me
300	5	2.3	77	1	Y4D1_RHISN	P55409 rhizobium s	373	5	2.3	114	1	EN78_HUMAN	P42830 homo sapien
301	5	2.3	79	1	NUAM_SIMVI	P50660 simulium vi	374	5	2.3	116	1	YJ90_YEAST	P47157 saccharomyc
302	5	2.3	80	1	COXM_HUMAN	P24311 homo sapien	375	5	2.3	116	1	CHH3_PENJP	Q94676 penaeus jap
303	5	2.3	80	1	COXM_MOUSE	P56393 mus musculu	376	5	2.3	117	1	PAND_HELPY	P56065 helicobacte
304	5	2.3	80	1	ICTA_SNY3	P27372 synecocyst	377	5	2.3	117	1	RS25_CAEEL	P52821 caenorhabdi
305	5	2.3	80	1	ITF_HUMAN	Q07654 homo sapien	378	5	2.3	117	1	B2MG_HORSE	P30441 equus cabal
306	5	2.3	81	1	UL11_VZVD	P09297 varicella-z	379	5	2.3	118	1	IAP4_NPVOP	Q10345 orgyia pseu
307	5	2.3	81	1	CHPR_ECOLI	P18534 escherichia	380	5	2.3	118	1	RL18_HELPY	Q92357 helicobacte
308	5	2.3	83	1	YK76_MYCTU	Q10684 mycobacteri	381	5	2.3	118	1	FOLB_STRCO	Q9x910 streptomyc
309	5	2.3	84	1	RL17_ODOSI	P49504 odontella s	382	5	2.3	119	1	RL18_HELPY	P56043 helicobacte
310	5	2.3	84	1	YOE8_ECOLI	P56605 escherichia	383	5	2.3	119	1	FOLB_BACSU	P28823 bacillus su
311	5	2.3	87	1	ECTP_TRIGR	P23227 tripeustes	384	5	2.3	120	1	ML1B_RAT	P49287 rattus norv
312	5	2.3	88	1	COXM_BOVIN	P13183 bos taurus	385	5	2.3	120	1	RL18_BACSU	P46899 bacillus su
313	5	2.3	89	1	GRLA_LEGPN	Q48833 legionella	386	5	2.3	120	1	Y061_BPT4	P13316 bacterioph
314	5	2.3	89	1	RS15_TREPA	Q83857 treponema p	387	5	2.3	120	1	NUOA_RHIME	O68952 rhizobium m
315	5	2.3	91	1	SY05_MOUSE	P30882 mus musculu	388	5	2.3	121	1	RECA_RUMAL	P49985 ruminooccu
316	5	2.3	92	1	SY05_RAT	P50231 rattus norv	389	5	2.3	121	1	RL17_CHLTR	P47760 chlamydia t
317	5	2.3	94	1	EM2_WHEAT	P22701 triticum ae	390	5	2.3	122	1	GLN2_METBA	P54809 methanosarc
318	5	2.3	94	1	FE45_BRAJA	Q45218 bradyrhizob	391	5	2.3	123	1	MSS4_HUMAN	P47224 homo sapien
319	5	2.3	94	1	IHPF_HAEIN	P43724 haemophilus	392	5	2.3	123	1	MSS4_RAT	Q08326 rattus norv
320	5	2.3	94	1	YOV3_CAEEL	P34661 caenorhabdi	393	5	2.3	123	1	RNPA_STRCO	P48206 streptomyc
321	5	2.3	95	1	CILG_HAEIN	P44461 haemophilus	394	5	2.3	124	1	GLU1_LOPAM	P01278 lophius ame
322	5	2.3	96	1	APC3_PIG	P27917 sus scrofa	395	5	2.3	124	1	YM06_MYCTU	Q10395 mycobacteri
323	5	2.3	96	1	GRO_RAT	P14095 rattus norv	396	5	2.3	125	1	RS11_AQUAE	O66485 aquifex aeo
324	5	2.3	96	1	RR6_GUITH	O78447 guillardia	397	5	2.3	125	1	RT11_WARPO	P26670 marchantia
325	5	2.3	96	1	YFRC_PROVU	P20927 proteus-vul	398	5	2.3	125	1		

399	1	125	5	2.3	1	P46744	prototetha	472	5	2.3	143	1	HS11_MEDSA	P27879	medicago sa
400	1	126	5	2.3	1	P25843	grosophylla	473	5	2.3	143	1	HVIC_HUMAN	P01744	homo sapien
401	1	127	5	2.3	1	P01545	hordeum vul	474	5	2.3	143	1	RNH_HELPJ	Q921h3	helicobacte
402	1	128	5	2.3	1	P42880	rhizobium m	475	5	2.3	143	1	RNH_HELPJ	P56120	helicobacte
403	1	129	5	2.3	1	Q22000	mycobacteri	476	5	2.3	143	1	RR11_MAIZE	P09561	zea mays (m
404	1	130	5	2.3	1	Q45791	bacteroides	477	5	2.3	143	1	RR11_MAIZE	P12096	oryza sativ
405	1	131	5	2.3	1	P27419	euglena gra	478	5	2.3	143	1	YK11_MCTU	Q10846	mycobacteri
406	1	132	5	2.3	1	P34403	caenorhabdi	479	5	2.3	144	1	FERI_LYCES	Q10846	mycobacteri
407	1	133	5	2.3	1	P40728	salmonella	480	5	2.3	144	1	FERI_LYCES	Q43517	lycopersico
408	1	134	5	2.3	1	P17647	cavia porce	481	5	2.3	144	1	NAPA_HELPJ	P43313	helicobacte
409	1	135	5	2.3	1	O78409	guillardia	482	5	2.3	144	1	NEU2_CAVPO	P10769	cavia porce
410	1	136	5	2.3	1	P06364	marchantia	483	5	2.3	145	1	YIHZ_HAEIN	P44814	haemophilus
411	1	137	5	2.3	1	P41630	pinus thunb	484	5	2.3	145	1	CYTE_HUMAN	O76096	homo sapien
412	1	138	5	2.3	1	O24709	synecococc	485	5	2.3	145	1	DHSD_CAEEL	O62215	caenorhabdi
413	1	139	5	2.3	1	P32988	synecocyst	486	5	2.3	145	1	Y004_TREPA	O83050	treponema p
414	1	140	5	2.3	1	Q00189	escherichia	487	5	2.3	145	1	YIHZ_ECOLI	P32147	escherichia
415	1	141	5	2.3	1	P04969	bacillus su	488	5	2.3	146	1	HBB1_XENTR	P07429	xenopus tro
416	1	142	5	2.3	1	P54218	caenorhabdi	489	5	2.3	146	1	OB_GORGO	O95189	gorilla gor
417	1	143	5	2.3	1	Q05806	tritricum ae	490	5	2.3	146	1	OB_PANTR	O02750	pan troglod
418	1	144	5	2.3	1	P39713	oncorhynch	491	5	2.3	147	1	GS17_XENLA	P07733	xenopus lae
419	1	145	5	2.3	1	P33745	oncorhynch	492	5	2.3	147	1	IL4_MESAU	O60440	mesocricetu
420	1	146	5	2.3	1	P17640	oncorhynch	493	5	2.3	147	1	IL4_MESAU	P20096	rattus norv
421	1	147	5	2.3	1	P19714	oncorhynch	494	5	2.3	147	1	MCP_MEDSA	P42067	medicago sa
422	1	148	5	2.3	1	P49648	odontella s	495	5	2.3	147	1	MUTX_BACFI	P26947	bacillus fi
423	1	149	5	2.3	1	P50303	felis silve	496	5	2.3	147	1	RS9_AQRAE	O67723	aquifex aeo
424	1	150	5	2.3	1	Q55625	synecocyst	497	5	2.3	147	1	YFP7_CAPVK	P19746	capripoxvir
425	1	151	5	2.3	1	O33752	synecococc	498	5	2.3	147	1	YI22_BURCE	P24537	burkholderi
426	1	152	5	2.3	1	P28910	escherichia	499	5	2.3	148	1	CPC4_CANPG	P81583	cancer pagu
427	1	153	5	2.3	1	Q00247	plectonema	500	5	2.3	148	1	Y4AN_RHISN	P55361	rhizobium s
428	1	154	5	2.3	1	O50467	neisseria g	501	5	2.3	149	1	TSJT_TOBAC	P24805	nicotiana t
429	1	155	5	2.3	1	P56364	chlorella v	502	5	2.3	149	1	YML3_THIFE	P20089	thiobacillu
430	1	156	5	2.3	1	O46901	guillardia	503	5	2.3	150	1	FERI_MAIZE	P27787	zea mays (m
431	1	157	5	2.3	1	P27789	zea mays (m	504	5	2.3	150	1	HS11_ORISA	P27777	oryza sativ
432	1	158	5	2.3	1	P30367	bos taurus	505	5	2.3	151	1	CH60_MYCAV	Q48900	mycobacteri
433	1	159	5	2.3	1	P21512	euglena gra	506	5	2.3	151	1	CH60_MYCCH	Q49093	mycobacteri
434	1	160	5	2.3	1	P16633	gracillaria	507	5	2.3	151	1	CH60_MYCFO	Q49160	mycobacteri
435	1	161	5	2.3	1	P27263	tomato yell	508	5	2.3	151	1	CH60_MYCMR	O50220	mycobacteri
436	1	162	5	2.3	1	P38610	tomato yell	509	5	2.3	151	1	HS11_WHEAT	P12810	tritricum ae
437	1	163	5	2.3	1	Q48835	legionella	510	5	2.3	151	1	RL19_METJA	P54043	methanococc
438	1	164	5	2.3	1	P38706	saccharomyc	511	5	2.3	151	1	RS13_PEA	P46298	pisum sativ
439	1	165	5	2.3	1	P30059	epifagus vi	512	5	2.3	151	1	VG55_BPMD2	O64246	mycobacteri
440	1	166	5	2.3	1	Q9Y4V0	aeropyrum p	513	5	2.3	151	1	YD88_SCHPO	O14183	schizosacch
441	1	167	5	2.3	1	O51313	borrelia bu	514	5	2.3	151	1	YK2A_YEAST	P36108	saccharomyc
442	1	168	5	2.3	1	P32032	tritricum ae	515	5	2.3	152	1	ANF_RAT	P01161	rattus norv
443	1	169	5	2.3	1	P21742	hordeum vul	516	5	2.3	152	1	YFF7_YEAST	P43346	saccharomyc
444	1	170	5	2.3	1	P01543	tritricum ae	517	5	2.3	153	1	HS11_HELAN	P30693	helianthus
445	1	171	5	2.3	1	P55886	salmonella	518	5	2.3	153	1	HS11_SOYBN	P02519	glycine max
446	1	172	5	2.3	1	P54706	dictyosteli	519	5	2.3	153	1	IP2K_SOLTU	P04793	glycine max
447	1	173	5	2.3	1	P29945	pseudomonas	520	5	2.3	153	1	VE6_MMPV	P01080	solanum tub
448	1	174	5	2.3	1	P29373	pseudomonas	521	5	2.3	153	1	VE6_MMPV	P07832	rattus norv
449	1	175	5	2.3	1	P22935	mus musculu	522	5	2.3	154	1	FASC_ECOLI	P30734	micromys mi
450	1	176	5	2.3	1	P08772	hordeum vul	523	5	2.3	154	1	HS12_ORISA	P45999	escherichia
451	1	177	5	2.3	1	P09618	hordeum vul	524	5	2.3	154	1	HS14_SOYBN	P31673	oryza sativ
452	1	178	5	2.3	1	P07854	spinacia ol	525	5	2.3	154	1	IP25_SOLTU	P04794	glycine max
453	1	179	5	2.3	1	Q29559	sus scrofa	526	5	2.3	154	1	RL30_HALMA	Q41488	solanum tub
454	1	180	5	2.3	1	P51673	rattus norv	527	5	2.3	155	1	VE6_HPV66	P14121	haloarcu
455	1	181	5	2.3	1	P41101	solanum tub	528	5	2.3	155	1	YPH7_CHRVI	Q80955	human papil
456	1	182	5	2.3	1	P48955	chondrus cr	529	5	2.3	155	1	YPH7_CHRVI	P45371	chromatium
457	1	183	5	2.3	1	P48955	chondrus cr	530	5	2.3	156	1	HS12_ARATH	P19036	arabidopsis
458	1	184	5	2.3	1	P51307	porphyra pu	531	5	2.3	156	1	SOXR_PSEAE	P13853	arabidopsis
459	1	185	5	2.3	1	Q924W4	rickettsia	532	5	2.3	156	1	HS11_DAUCA	Q51506	pseudomonas
460	1	186	5	2.3	1	P06587	pisum sativ	533	5	2.3	157	1	YI2F_BACSU	P27396	daucus caro
461	1	187	5	2.3	1	P06365	nicotiana t	534	5	2.3	157	1	DHSD_BOVIN	O32068	bacillus su
462	1	188	5	2.3	1	P16795	human cytom	535	5	2.3	158	1	HPPK_METEX	Q95123	bos taurus
463	1	189	5	2.3	1	P50807	bovine papi	536	5	2.3	158	1	HS11_PEA	P71512	methylobact
464	1	190	5	2.3	1	Q14695	homo sapien	537	5	2.3	158	1	YBHB_ECOLI	P19243	pisum sativ
465	1	191	5	2.3	1	P51307	porphyra pu	538	5	2.3	158	1	YBHB_ECOLI	P12994	escherichia
466	1	192	5	2.3	1	Q924W4	rickettsia	539	5	2.3	158	1	YBHB_ECOLI	P12994	escherichia
467	1	193	5	2.3	1	P03795	bacterioph	540	5	2.3	159	1	19KD_MCTU	O05515	bacillus su
468	1	194	5	2.3	1	P14362	fovilpox vir	541	5	2.3	159	1	MP63_MCTU	P11572	mycobacteri
469	1	195	5	2.3	1	Q14377	schizosacch	542	5	2.3	159	1	SMPA_TREHY	P9175	mycobacteri
470	1	196	5	2.3	1	P43718	haemophilus	543	5	2.3	160	1	RS4_SULAC	Q54313	treponema h
471	1	197	5	2.3	1	P41055	streptomyce	544	5	2.3	161	1	19KD_MYCAV	P39467	sulfolobus
	1	198	5	2.3	1	P39611	bacillus su		5	2.3	161	1	HS11_CHERU	P46733	mycobacteri
	1	199	5	2.3	1				5	2.3	161	1		Q05832	chenopodium

545	161	1	2.3	5	19037	arabidopsis	618	5	2.3	180	1	Y52L_PROMA	Q51893	prochloroco
546	161	1	2.3	5	P05478	glycine max	619	5	2.3	180	1	YMER_STAAU	P08655	staphylococ
547	161	1	2.3	5	P53922	saccharomyc	620	5	2.3	180	1	YRP2_RHET	O69777	rhizobium e
548	162	1	2.3	5	P31502	mycobacteri	621	5	2.3	181	1	ARF2_MOUSE	P16500	mus musculus
549	162	1	2.3	5	P48263	cyanophora	622	5	2.3	181	1	OM20_NEUCR	P35848	neurospora
550	162	1	2.3	5	O78454	guillardia	623	5	2.3	181	1	RS4_SULSO	P95987	sulfolobus
551	162	1	2.3	5	Q60286	methanococc	624	5	2.3	182	1	ARF_DUGJA	P91924	dugesia jap
552	163	1	2.3	5	P49510	odontella s	625	5	2.3	182	1	NOO9_THETH	Q56224	thermus agu
553	163	1	2.3	5	P51199	porphyra pu	626	5	2.3	182	1	SECY_PARDE	P72179	paracoccus
554	163	1	2.3	5	P45260	haemophilus	627	5	2.3	182	1	Y922_HAEIN	Q57457	haemophilus
555	163	1	2.3	5	P29921	paracoccus	628	5	2.3	182	1	YLP3_CAEEL	P34383	caenorhabdi
556	163	1	2.3	5	P13333	bacterioph	629	5	2.3	183	1	IR14_RCMVA	P21601	human cytom
557	163	1	2.3	5	Q38435	bacterioph	630	5	2.3	183	1	YI71_BURCE	P24577	chlamydia p
558	164	1	2.3	5	O57608	phasianus c	631	5	2.3	184	1	GRPE_CHLPN	Q92849	burkholderi
559	164	1	2.3	5	P28999	synecococc	632	5	2.3	184	1	NUSG_BORBU	O51355	borrelia bu
560	164	1	2.3	5	O69583	mycobacteri	633	5	2.3	184	1	TRSF_DR OSI	Q24366	drosophila
561	164	1	2.3	5	P34477	caenorhabdi	634	5	2.3	184	1	Y487_MYCLE	P54139	mycobacteri
562	164	1	2.3	5	P76510	escherichia	635	5	2.3	185	1	HID_STRPU	P15870	strongyloce
563	165	1	2.3	5	P12061	salmonella	636	5	2.3	185	1	J1PH_ATRCA	P42764	atriplex ca
564	165	1	2.3	5	P55670	rhizobium s	637	5	2.3	185	1	PSAF_PORPU	P51193	porphyra pu
565	167	1	2.3	5	Q01545	pharbitis n	638	5	2.3	185	1	SODC_FRATU	Q59448	francisella
566	167	1	2.3	5	P41159	homo sapien	639	5	2.3	185	1	YEP6_YEAST	P35750	sus scrofa
567	167	1	2.3	5	P42747	arabidopsis	640	5	2.3	186	1	CAN1_PIG	P15454	saccharomyc
568	168	1	2.3	5	P35434	rattus norv	641	5	2.3	186	1	KGUA_YEAST	Q08677	photobacter
569	168	1	2.3	5	Q95136	bos taurus	642	5	2.3	186	1	LOXP_PHOLE	P73376	synecocyst
570	168	1	2.3	5	P14448	human immun	643	5	2.3	186	1	Y50L_SYNY3	Q56416	thermus agu
571	168	1	2.3	5	P12204	nicotiana t	644	5	2.3	186	1	YACE_THETH	O08545	mus musculus
572	169	1	2.3	5	O78451	guillardia	645	5	2.3	187	1	EFA3_MOUSE	Q9x2h4	thermotoga
573	169	1	2.3	5	Q46578	deinococcus	646	5	2.3	187	1	Y160_THEMA	O9x2dt7	rickettsia
574	169	1	2.3	5	Q9x2dt7	rickettsia	647	5	2.3	188	1	FFP_RICPR	P15950	rattus norv
575	169	1	2.3	5	P52806	pinus thunb	648	5	2.3	188	1	KLK3_RAT	P23593	rattus norv
576	169	1	2.3	5	Q00644	azospirillum	649	5	2.3	189	1	APD_RAT	P10463	canis fami
577	169	1	2.3	5	P13116	escherichia	650	5	2.3	189	1	CAYP_CANFA	P41150	cryptotolagu
578	170	1	2.3	5	P24515	drosophila	651	5	2.3	189	1	CAYP_RABIT	P30583	emeritella
579	170	1	2.3	5	Q99462	homo sapien	652	5	2.3	189	1	CHS1_EMENI	P49878	bos taurus
580	170	1	2.3	5	P27324	zea mays (m	653	5	2.3	189	1	INAH_BOVIN	P28915	escherichia
581	170	1	2.3	5	YCF3_ORYSA	oryza sativ	654	5	2.3	189	1	YBFC_ECOLI	P77717	escherichia
582	171	1	2.3	5	LSFA_HAEIN	haemophilus	655	5	2.3	190	1	YBAY_ECOLI	Q20742	schizosacch
583	172	1	2.3	5	ILVH_SYNY3	synecocyst	656	5	2.3	191	1	RR4_CYAME	O09290	cyanidiosch
584	172	1	2.3	5	Y21B_METJA	methanococc	657	5	2.3	191	1	YB61_SCHPO	Q20742	schizosacch
585	173	1	2.3	5	PRSE_ECOLI	escherichia	658	5	2.3	192	1	HS41_SOYBN	P30236	glycine max
586	174	1	2.3	5	CALB_YEAST	saccharomyc	659	5	2.3	192	1	RAC2_HUMAN	P15153	homo sapien
587	174	1	2.3	5	DRDR_BOVIN	bos taurus	660	5	2.3	192	1	RAC2_MOUSE	Q05144	mus musculus
588	174	1	2.3	5	DSEC_SALTY	salmonella	661	5	2.3	192	1	RR4_CYACA	O22029	cyanidium c
589	174	1	2.3	5	FMA_SERMA	serratia ma	662	5	2.3	193	1	HS72_CANAL	P46587	candida alb
590	174	1	2.3	5	GSPB_KLEPN	klebsiella	663	5	2.3	193	1	PAPJ_ECOLI	P17543	escherichia
591	174	1	2.3	5	ILVH_PORPU	porphyra pu	664	5	2.3	193	1	PRSJ_ECOLI	P42189	escherichia
592	174	1	2.3	5	SSBP_ECOLI	escherichia	665	5	2.3	193	1	RNH2_RICPR	Q9x2dw3	rickettsia
593	175	1	2.3	5	FLAW_KLEPN	escherichia	666	5	2.3	193	1	YB71_HAEIN	P44339	haemophilus
594	175	1	2.3	5	IPYR_ECOLI	escherichia	667	5	2.3	193	1	YQFW_BACSU	P4480	bacillus su
595	175	1	2.3	5	YCB8_PSEDE	pseudomonas	668	5	2.3	194	1	ACPD_HAEIN	P43013	haemophilus
596	176	1	2.3	5	LALP_MACEU	macropus eu	669	5	2.3	195	1	HPPK_SYNY3	P72736	synecocyst
597	176	1	2.3	5	LALP_TRIUV	trichosurus	670	5	2.3	195	1	MOG_ECOLI	P28694	escherichia
598	176	1	2.3	5	RL6_ECOLI	escherichia	671	5	2.3	195	1	RS7_SCHPO	Q10101	schizosacch
599	176	1	2.3	5	V19R_VACCV	vaccinia vi	672	5	2.3	196	1	PSAD_CHLRE	Q39615	chlamydomon
600	177	1	2.3	5	RL6_ACYKS	acrythosiph	673	5	2.3	196	1	RNH2_AQUAE	P4480	baquifex aeo
601	177	1	2.3	5	MYCB_RAT	rattus norv	674	5	2.3	196	1	THIJ_ECOLI	Q46948	escherichia
602	178	1	2.3	5	RL6_BACSU	baquifex aeo	675	5	2.3	196	1	TRAP_ECOLI	P41068	escherichia
603	178	1	2.3	5	RL6_HELPU	haemophilus	676	5	2.3	197	1	HS41_PEA	P15244	pisum sativ
604	178	1	2.3	5	RL6_HELPU	haemophilus	677	5	2.3	197	1	MOG_HAEIN	P44645	haemophilus
605	178	1	2.3	5	RPOB_GDAB	gallinacea	678	5	2.3	197	1	TRSF_DRONE	P11596	drosophila
606	178	1	2.3	5	SSBF_ECOLI	escherichia	679	5	2.3	197	1	XYNA_SCHRO	P35809	schizophyll
607	179	1	2.3	5	PTH_CHLTR	chlamydia t	680	5	2.3	197	1	Y882_METJA	Q58292	methanococc
608	179	1	2.3	5	SP17_MONDO	monodelphis	681	5	2.3	197	1	NGAL_HUMAN	Q58338	methanococc
609	179	1	2.3	5	TRBB_ECOLI	trichosurus	682	5	2.3	198	1	NDC_EMENI	P80188	homo sapien
610	180	1	2.3	5	APT_STOLO	apicomplexa	683	5	2.3	198	1	YAEH_HAEIN	P17624	emeritella
611	180	1	2.3	5	ARF1_HUMAN	arabidopsis	684	5	2.3	198	1	AA27_HUMAN	P46492	haemophilus
612	180	1	2.3	5	ARF1_XENLA	xenopus lae	685	5	2.3	199	1	IL11_HUMAN	O60232	homo sapien
613	180	1	2.3	5	ARF3_HUMAN	arabidopsis	686	5	2.3	199	1	IL11_MOUSE	P20809	homo sapien
614	180	1	2.3	5	DADR_RABIT	dactylocten	687	5	2.3	199	1	NU6M_MARPO	P47873	mus musculus
615	180	1	2.3	5	GLUC_BOVIN	glucosylated	688	5	2.3	199	1	NUOC_RHOCA	P26850	marichantia
616	180	1	2.3	5	GLUC_BOVIN	glucosylated	689	5	2.3	199	1	TXD2_BRUMA	O84971	rhodobacter
617	180	1	2.3	5	GLUC_MESAU	mesocricetu	690	5	2.3	199	1		Q17172	brugia mala

691	2.3	200	1	NGAL_MOUSE	P11672	mus musculus	764	5	2.3	217	1	YQIE_HAEIN	P44684	haemophilus
692	2.3	200	1	RS8_SCHPO	O14049	schizosacch	765	5	2.3	218	1	GTW4_HUMAN	Q03013	homo sapien
693	2.3	200	1	STRM_STRGR	P29783	streptomyce	766	5	2.3	218	1	KAD_BORPE	P39068	bordetella
694	2.3	200	1	YOR1_COYMY	P19200	commelina y	767	5	2.3	218	1	RR3_PICAB	O62951	picea abies
695	2.3	201	1	BLIP_STRCL	P35804	streptomyce	768	5	2.3	218	1	RS7_PYRHO	O59230	pyrococcus
696	2.3	201	1	YXBE_BACSU	P37872	bacillus su	769	5	2.3	218	1	Y144_CAEBL	Q11072	caenorhabdi
697	2.3	202	1	CUTI_BTCCI	Q00298	botrytis ci	770	5	2.3	219	1	AROD_AQUAE	O66440	aquifex aeo
698	2.3	202	1	ETI_BOVIN	P17322	bos taurus	771	5	2.3	219	1	PIGF_HUMAN	Q07326	homo sapien
699	2.3	202	1	ET1_RAT	P22388	rattus norv	772	5	2.3	219	1	YJFK_ECOLI	P39293	escherichia
700	2.3	202	1	HS2M_PEA	P46254	pisum sativ	773	5	2.3	219	1	YLP4_ZYMO	O66114	zymomonas m
701	2.3	202	1	RB18_LYAST	O05376	lymaea sta	774	5	2.3	220	1	GTCL_RAT	P04904	rattus norv
702	2.3	202	1	IY72_YEAST	P47142	saccharomyc	775	5	2.3	220	1	GTG2_RAT	P46418	rattus norv
703	2.3	203	1	CAIE_ECOLI	P39206	escherichia	776	5	2.3	220	1	GTG_MOUSE	P30115	mus musculus
704	2.3	203	1	COBC_ECOLI	P52086	escherichia	777	5	2.3	220	1	NK14_MOUSE	P27814	mus musculus
705	2.3	203	1	GTS_OMSL	P46088	ommastrephe	778	5	2.3	220	1	PC83_SOLTU	P20347	solanum tub
706	2.3	203	1	RAB7_DICDI	P36411	dictyosteli	779	5	2.3	220	1	RNH2_BRUME	O68821	bruceella me
707	2.3	204	1	TRPG_HALVO	P33974	halobacteri	780	5	2.3	220	1	RNH2_PYRHO	O59351	pyrococcus
708	2.3	205	1	DUT_RAT	P70583	rattus norv	781	5	2.3	221	1	ATP6_CLOAB	O05097	clostridium
709	2.3	205	1	RCP_VIBCH	P24550	vibrio chol	782	5	2.3	221	1	GTAL_PIG	P51781	sus scrofa
710	2.3	205	1	RS4_MYCPE	P47553	mycoplasma	783	5	2.3	222	1	KADC_MAIZE	P43188	zea mays (m
711	2.3	205	1	RS4_MYCPN	P46775	mycoplasma	784	5	2.3	222	1	UL20_HSV11	P10204	herpes simp
712	2.3	205	1	Y115_NPVOP	O10354	argyria pseu	785	5	2.3	222	1	Y4YL_RHISN	P55720	rhizobium s
713	2.3	205	1	YETR_SHEPU	P46149	shevanella	786	5	2.3	222	1	YKGH_ECOLI	P77180	escherichia
714	2.3	206	1	R13A_PICMA	O65055	picea maria	787	5	2.3	223	1	B1OD_MYCLE	P45486	mycobacteri
715	2.3	206	1	RL3_THEFH	P52860	thermus aqu	788	5	2.3	223	1	CYPH_NEUCR	P10255	neurospora
716	2.3	206	1	RNH2_METH	O27102	methanobact	789	5	2.3	223	1	NARI_BACSU	P42177	bacillus su
717	2.3	206	1	TRIC_RABIT	P02646	oryctolagus	790	5	2.3	224	1	PLF1_MOUSE	P04095	mus musculus
718	2.3	207	1	HIS7_AZOB	P18787	azospirillu	791	5	2.3	224	1	PLF2_MOUSE	P04768	mus musculus
719	2.3	207	1	NU6M_PROWI	Q37626	protobeca	792	5	2.3	224	1	PLF3_MOUSE	P18918	mus musculus
720	2.3	207	1	UPP_MYCTU	P94928	mycobacteri	793	5	2.3	224	1	UL01_HCMVA	P16719	human cytom
721	2.3	208	1	HS26_DROME	P02517	drosophila	794	5	2.3	224	1	UREG_MYCTU	P50051	mycobacteri
722	2.3	208	1	KGUA_MYCTU	P71659	mycobacteri	795	5	2.3	224	1	Y119_METJA	Q57583	methanococc
723	2.3	208	1	NU4M_MICPE	O21521	microtus pe	796	5	2.3	225	1	KDPE_ECOLI	P21866	escherichia
724	2.3	208	1	PSAD_LYCES	P12372	lycopersico	797	5	2.3	225	1	OLIG_HUMAN	P47887	homo sapien
725	2.3	208	1	RL13_CHLSW	O48513	chlamydomon	798	5	2.3	226	1	BASP_BOVIN	P80724	bos taurus
726	2.3	209	1	PGSA_MYCTU	O50611	mycobacteri	799	5	2.3	226	1	C79A_HUMAN	P11912	homo sapien
727	2.3	209	1	RNH2_HELPJ	Q92111	helicobacte	800	5	2.3	226	1	LYCV_BPPHV	P13003	bacterioph
728	2.3	209	1	RNH2_HELPJ	P56121	helicobacte	801	5	2.3	226	1	YC65_MYCTO	Q11054	mycobacteri
729	2.3	209	1	YRKB_HAEIN	P45075	haemophilus	802	5	2.3	227	1	DEFA_ANTMA	P23706	antirrhinum
730	2.3	210	1	CUTI_PHYCP	P41754	phytophthor	803	5	2.3	227	1	EXOD_BPT4	P04536	bacterioph
731	2.3	210	1	ESAI_ERWST	P54656	erwinia ste	804	5	2.3	228	1	COX2_LUMTE	Q37545	lumbritic t
732	2.3	210	1	Y2G1_YEAST	P53251	saccharomyc	805	5	2.3	228	1	LPRH_MYCTU	P71687	mycobacteri
733	2.3	211	1	CCMA_PARDE	P52218	paracoccus	806	5	2.3	228	1	MENE_SALTY	P37418	salmonella
734	2.3	211	1	PMGY_SCHPO	P55165	gallus gall	807	5	2.3	228	1	MP64_MYCTU	P19996	mycobacteri
735	2.3	211	1	RNB3_CHICK	P36623	schizosacch	808	5	2.3	228	1	RNH2_PYRKO	O74035	pyrococcus
736	2.3	211	1	RCSA_ERWST	P20098	erwinia amy	809	5	2.3	228	1	YBFG_HAEIN	P46670	haemophilus
737	2.3	211	1	RCSA_ERWST	P27488	erwinia ste	810	5	2.3	229	1	FLGF_RHIME	O54248	rhizobium m
738	2.3	211	1	RS3A_ARCFU	O27964	archaeoglob	811	5	2.3	229	1	NHB1_RHORH	P21220	rhodococcus
739	2.3	211	1	TER5_ECOLI	P21337	escherichia	812	5	2.3	229	1	REPX_STAUA	P03862	staphylococ
740	2.3	212	1	PCPP_BPT4	P06807	bacterioph	813	5	2.3	229	1	VG01_BPML5	Q05218	mycobacteri
741	2.3	212	1	RL7_FICJA	P32101	pichia jadl	814	5	2.3	229	1	XYN1_TREIRE	P36218	trichoderma
742	2.3	213	1	GPH_AQUAE	O67359	aquifex aeo	815	5	2.3	230	1	CRP_RAT	P48199	rattus norv
743	2.3	213	1	UBC_ASFM2	P25869	african swi	816	5	2.3	230	1	D102_ARATH	Q05212	arabidopsis
744	2.3	213	1	XJH2_YEAST	P40359	saccharomyc	817	5	2.3	230	1	R1BR_BACSU	P94465	bacillus su
745	2.3	214	1	R1R2_DROME	P48592	drosophila	818	5	2.3	230	1	RK1_ODOSI	P49544	odontella s
746	2.3	214	1	RNH2_CHLPN	Q92962	chlamydia p	819	5	2.3	230	1	RNH2_METJA	O57599	methanococc
747	2.3	214	1	Y629_METJA	Q58046	methanococc	820	5	2.3	230	1	SOML_HPHPI	P45641	hippoglossu
748	2.3	214	1	Y706_METJA	Q58117	methanococc	821	5	2.3	230	1	SOML_SOLSE	P45642	solea seneg
749	2.3	214	1	Y023_CAEBL	P34673	caenorhabdi	822	5	2.3	231	1	COX2_ALOPA	P98024	alouatta pa
750	2.3	215	1	YPU5_RHOCA	P26161	rhodobacter	823	5	2.3	231	1	COX2_LAGLA	P98036	lagothrix l
751	2.3	215	1	UBC_ASFB7	P27949	african swi	824	5	2.3	231	1	MAD1_PETHY	Q07472	petunia hyb
752	2.3	216	1	COX2_CALGO	P50661	callimico g	825	5	2.3	231	1	NFIL_PIG	P14057	sus scrofa
753	2.3	216	1	OLIA_HUMAN	P47882	homo sapien	826	5	2.3	231	1	Y282_MYCTU	P96864	mycobacteri
754	2.3	216	1	OL1D_HUMAN	P47885	homo sapien	827	5	2.3	232	1	GUS6_RAT	P35899	rattus norv
755	2.3	217	1	GTM1_HUMAN	P09488	homo sapien	828	5	2.3	232	1	RSUA_HAEIN	P45124	haemophilus
756	2.3	217	1	GTM1_RAT	P04905	rattus norv	829	5	2.3	232	1	YACM_BACSU	Q06755	bacillus su
757	2.3	217	1	GTM2_HUMAN	P28161	homo sapien	830	5	2.3	232	1	YHCG_BACSU	P54591	bacillus su
758	2.3	217	1	GTM3_RAT	P08009	rattus norv	831	5	2.3	232	1	YHAK_ECOLI	P42624	escherichia
759	2.3	217	1	GTM5_HUMAN	P46439	homo sapien	832	5	2.3	233	1	YM27_MYCTU	Q10511	mycobacteri
760	2.3	217	1	GTMO_CAVPO	P16413	cavia porce	833	5	2.3	234	1	COBC_SALTY	P39701	salmonella
761	2.3	217	1	HNT2_YEAST	P49775	saccharomyc	834	5	2.3	234	1	DNAA_RHILE	Q33529	rhizobium l
762	2.3	217	1	RNH2_CHLTR	O84032	chlamydia t	835	5	2.3	234	1	Y014_SYNY3	P57208	synecocyst
763	2.3	217	1	RR3_PINTH	P41635	pinus thunb	836	5	2.3	235	1	Y544_RICPR	Q92G06	rickettsia

837	5	2.3	236	1	ARCA_HAEIN	P44918 haemophilus	910	5	2.3	253	1	ADH1_DROMU	P07161 drosophila
838	5	2.3	236	1	CLCD_PSEPU	P11453 pseudomonas	911	5	2.3	253	1	ADH1_DRONA	P12854 drosophila
839	5	2.3	236	1	G76_SCHMA	P46435 schistosoma	912	5	2.3	253	1	ADH2_DROBU	P25720 drosophila
840	5	2.3	236	1	RECO_HAEIN	P44642 haemophilus	913	5	2.3	253	1	ADH2_DROHY	P23237 drosophila
841	5	2.3	237	1	ATPF_KLUJA	O13349 kluyveromyc	914	5	2.3	253	1	ADH2_DROMU	P07160 drosophila
842	5	2.3	237	1	RECA_NEIGS	O86382 neisseria e	915	5	2.3	253	1	ADH2_DROMY	P25721 drosophila
843	5	2.3	237	1	RECA_NEILA	P97195 neisseria l	916	5	2.3	253	1	ADH2_DROWH	P24267 drosophila
844	5	2.3	237	1	RECA_NEIFE	O86403 neisseria p	917	5	2.3	253	1	ADH_DROAD	Q00669 drosophila
845	5	2.3	237	1	RECA_NEIPO	P96955 neisseria p	918	5	2.3	253	1	ADH_DRODI	P22245 drosophila
846	5	2.3	237	1	RECA_NEISI	O86411 neisseria s	919	5	2.3	253	1	ADH_DROGE	Q09009 drosophila
847	5	2.3	237	1	RECA_NEISU	O86414 neisseria s	920	5	2.3	253	1	ADH_DROHE	P21898 drosophila
848	5	2.3	237	1	Y576_CAREL	Q09619 caenorhabdi	921	5	2.3	253	1	ADH_DROIM	Q07588 drosophila
849	5	2.3	238	1	EA3_HUMAN	P52797 homo sapien	922	5	2.3	253	1	ADH_DROMD	Q09010 drosophila
850	5	2.3	238	1	HS2C_WHEAT	Q00445 triticum ae	923	5	2.3	253	1	ADH_DROPE	P37473 drosophila
851	5	2.3	238	1	PRRC_RAT	P33579 rattus norv	924	5	2.3	253	1	ADH_DROPL	P23277 drosophila
852	5	2.3	238	1	RNH2_THEMA	Q9X017 thermotoga	925	5	2.3	253	1	ADH_DROPS	P07158 drosophila
853	5	2.3	238	1	TCBE_PSEPO	P27100 pseudomonas	926	5	2.3	253	1	ADH_DROSL	P23278 drosophila
854	5	2.3	238	1	UBC3_RABIT	Q29503 oryctolagus	927	5	2.3	253	1	RM02_ACACA	P46763 acanthamoeb
855	5	2.3	238	1	Y554_AQUAE	O66829 aquifex aeo	928	5	2.3	253	1	SSRA_ARATH	P45434 arabidopsis
856	5	2.3	239	1	APBL_AERPE	O73942 aeropyrum p	929	5	2.3	253	1	YMDA_CHLAU	Q45826 chloroflexu
857	5	2.3	239	1	CLDE_HUMAN	O95500 homo sapien	930	5	2.3	254	1	ADH_DROAM	P25139 drosophila
858	5	2.3	239	1	CLDE_MOUSE	Q920s3 mus musculu	931	5	2.3	254	1	ADH_DROLE	P10807 drosophila
859	5	2.3	239	1	CYBH_RHILV	P27648 rhizobium l	932	5	2.3	254	1	ADH_DROSU	Q03384 drosophila
860	5	2.3	239	1	GUNA_ASPAK	Q12679 aspergillus	933	5	2.3	254	1	MOTA_AQUAE	O67122 aquifex aeo
861	5	2.3	239	1	KLK2_CAVPO	P12323 cavia porce	934	5	2.3	254	1	PRG6_YEAST	P40303 saccharomyc
862	5	2.3	239	1	PNUC_ECOLI	P31215 escherichia	935	5	2.3	254	1	YBFF_ECOLI	P75736 escherichia
863	5	2.3	239	1	RL2_PYRHO	O59421 pyrococcus	936	5	2.3	254	1	YH11_MYCTU	O33210 mycobacteri
864	5	2.3	239	1	UL20_HSVB	P28971 equine herp	937	5	2.3	254	1	YNO8_YEAST	P35904 saccharomyc
865	5	2.3	239	1	Y0IK_BACSU	P54527 bacillus su	938	5	2.3	255	1	143P_ARATH	P24644 arabidopsis
866	5	2.3	239	1	YVIC_BACSU	P46909 bacillus su	939	5	2.3	255	1	ADH_DROER	P28483 drosophila
867	5	2.3	240	1	CD48_MOUSE	P18181 mus musculu	940	5	2.3	255	1	ADH_DROMA	P07162 drosophila
868	5	2.3	240	1	RNC_MYCTU	Q10962 mycobacteri	941	5	2.3	255	1	ADH_DROME	P00334 drosophila
869	5	2.3	240	1	SURL_FUGRU	O57593 figu rubrip	942	5	2.3	255	1	ADH_DROSE	P07159 drosophila
870	5	2.3	240	1	YDGB_ECOLI	P52109 escherichia	943	5	2.3	255	1	ADH_DROSI	P07163 drosophila
871	5	2.3	241	1	BYR_BORPE	Q08530 bordetella	944	5	2.3	255	1	ADH_DROTE	P28484 drosophila
872	5	2.3	241	1	GLTL_ECOLI	P41076 escherichia	945	5	2.3	255	1	ADH_DROYA	P26719 drosophila
873	5	2.3	241	1	HUTC_KLEAE	P12380 klebsiella	946	5	2.3	255	1	CAPC_STANU	P39852 staphylococ
874	5	2.3	241	1	TRPF_METTH	O27695 methanobact	947	5	2.3	255	1	STX6_HUMAN	O43752 homo sapien
875	5	2.3	241	1	YD96_YEAST	Q03941 saccharomyc	948	5	2.3	255	1	STX6_RAT	Q63635 rattus norv
876	5	2.3	242	1	GLUA_CORGL	P48243 corynebacte	949	5	2.3	255	1	YHJH_ECOLI	P37646 escherichia
877	5	2.3	242	1	RECO_ECOLI	P15027 escherichia	950	5	2.3	256	1	ADH_DROTS	P51550 drosophila
878	5	2.3	242	1	RL2_METJA	P54017 methanococc	951	5	2.3	256	1	ATPF_HUMAN	P24539 homo sapien
879	5	2.3	243	1	AG16_TRYBB	Q26768 trypanosoma	952	5	2.3	256	1	ATPF_RAT	P19511 rattus norv
880	5	2.3	243	1	TRYL_XENLA	P19799 xenopus lae	953	5	2.3	256	1	PXBA_CANBO	Q00316 candida boi
881	5	2.3	243	1	YNO6_CAEEL	Q03615 caenorhabdi	954	5	2.3	256	1	PXBB_CANBO	O00317 candida boi
882	5	2.3	244	1	FMCH_BACNO	P17419 bacteroides	955	5	2.3	256	1	Y107_NPVOP	O10346 orgyia pseu
883	5	2.3	244	1	KLKA_RAT	P36375 rattus norv	956	5	2.3	256	1	YAFV_ECOLI	Q47679 escherichia
884	5	2.3	244	1	RDGA_ERWCA	Q47587 erwinia car	957	5	2.3	256	1	YH11_MYCLE	O05668 mycobacteri
885	5	2.3	244	1	UPP_TOXGO	Q26998 toxoplasma	958	5	2.3	257	1	ABCX_CYACA	P35020 cyanidium c
886	5	2.3	245	1	DCOP_SALTY	P07691 salmonella	959	5	2.3	257	1	KLK1_MACFA	Q07276 macaca fasc
887	5	2.3	245	1	LUXP_VIBHA	P54300 vibrio harv	960	5	2.3	257	1	NOCP_AGRTS	P35116 agrobacteri
888	5	2.3	245	1	PMX1_HUMAN	P54821 homo sapien	961	5	2.3	257	1	TXDM_MOUSE	P20108 mus musculu
889	5	2.3	246	1	PMX1_MOUSE	P43271 mus musculu	962	5	2.3	257	1	Y036_MYCTU	P71606 mycobacteri
890	5	2.3	246	1	WECG_ECOLI	P27836 escherichia	963	5	2.3	258	1	ARA5_ARATH	P28188 arabidopsis
891	5	2.3	246	1	Y181_METJA	Q57640 methanococc	964	5	2.3	258	1	CODY_BACSU	P39779 bacillus su
892	5	2.3	246	1	YP73_BPP2	Q50648 mycobacteri	965	5	2.3	258	1	KLK1_LAPHA	Q28773 papio hamad
893	5	2.3	247	1	VPM_BPP2	P25476 bacterioph	966	5	2.3	258	1	YJ98_MYCTU	Q10859 mycobacteri
894	5	2.3	248	1	HI_PARAN	P02256 parechinus	967	5	2.3	258	1	YWN8_YEAST	P53862 saccharomyc
895	5	2.3	248	1	PRC8_DICDI	Q27563 dictyosteli	968	5	2.3	259	1	CAH2_BOVIN	P00921 bos taurus
896	5	2.3	248	1	RSR1_CANAL	P52498 candida alb	969	5	2.3	259	1	CAH2_SHEEP	P00922 ovis aries
897	5	2.3	249	1	CDSA_ECOLI	P06466 escherichia	970	5	2.3	259	1	KLK2_RAT	P00759 rattus norv
898	5	2.3	249	1	HMX1_CHICK	P28361 gallus gall	971	5	2.3	259	1	KLK9_RAT	P07647 rattus norv
899	5	2.3	249	1	RU2A_ARATH	P43333 arabidopsis	972	5	2.3	259	1	KLKB_RAT	P35976 rattus norv
900	5	2.3	249	1	YE64_AQUAE	O67444 aquifex aeo	973	5	2.3	259	1	KLKL_MOUSE	P15948 mus musculu
901	5	2.3	249	1	YS91_MYCTU	Q10812 mycobacteri	974	5	2.3	259	1	NODJ_RHIV	P05755 rhizobium l
902	5	2.3	250	1	ODEP_HUMAN	Q14990 homo sapien	975	5	2.3	259	1	PBBI_CAMJE	P45678 campylobact
903	5	2.3	251	1	GLO2_ECOLI	Q47677 escherichia	976	5	2.3	259	1	RNPH_MYCLE	P37939 mycobacteri
904	5	2.3	251	1	TPIS_LEIME	P48499 leishmania	977	5	2.3	259	1	SSAT_SALTY	P69068 salmonella
905	5	2.3	252	1	AG_BRANA	Q01540 brassica na	978	5	2.3	259	1	YBBM_ECOLI	P77307 escherichia
906	5	2.3	252	1	P29_MYCHR	P15361 mycoplasma	979	5	2.3	260	1	ARGT_SALTY	P02911 salmonella
907	5	2.3	253	1	ADH1_DROHY	P23236 drosophila	980	5	2.3	260	1	CLCA_PSEPU	P11451 pseudomonas
908	5	2.3	253	1	ADH1_DROMO	P09370 drosophila	981	5	2.3	260	1	COX3_PISOC	P25003 pisaster oc
909	5	2.3	253	1	ADH1_DROMT	P22246 drosophila	982	5	2.3	260	1	NMA_HUMAN	Q13145 homo sapien

983 5 2.3 260 1 RFXK_HUMAN
984 5 2.3 261 1 CATC_MOUSE
985 5 2.3 261 1 COX3_HUMAN
986 5 2.3 261 1 KKK1_MOUSE
987 5 2.3 261 1 KKK3_MOUSE
988 5 2.3 261 1 KKK6_MOUSE
989 5 2.3 261 1 KKK7_MOUSE
990 5 2.3 261 1 KKK8_MOUSE
991 5 2.3 261 1 KKK9_MOUSE
992 5 2.3 261 1 KKKF_MOUSE
993 5 2.3 261 1 MOTB_BACSU
994 5 2.3 261 1 PRC9_HUMAN
995 5 2.3 261 1 PRC9_MOUSE
996 5 2.3 261 1 PROS_MACMU
997 5 2.3 261 1 UREH_HAEIN
998 5 2.3 262 1 KKK1_HUMAN
999 5 2.3 262 1 ODFP_BOVIN
1000 5 2.3 262 1 ODFP_PIG

ALIGNMENTS

RESULT 1
YMFEC_ECOLI STANDARD; PRT; 217 AA.
AC P75966;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 24.9 KD PROTEIN IN TRMU-ICDA INTERGENIC REGION.
GN YMFEC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 97061202.
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map";
RL DNA Res. 3:137-155(1996).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC STRONG, TO H. INFLUENZAE HI0694.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000213; AAC74219.1; ALT_INIT.
CC EMBL; D90748; BAA35957.1; -
CC EMBL; D90749; BAA35966.1; -
CC ECOGENE; EG13447; YMFEC.

DR PROSITE; PS01149; PSI_RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 217 AA; 24880 MW; F7C7A7CEDC5FD3F6 CRC64;

Query Match 23.4%; Score 52; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 WLMPRPPIREKRSIPTSLKITYEGRNRQVRMTAHTVGFPTLRIRYAMG 202
|||||
DB 149 WLMPRPPIREKRSIPTSLKITYEGRNRQVRMTAHTVGFPTLRIRYAMG 200
|||||

RESULT 2
YMFEC_HAEIN STANDARD; PRT; 240 AA.
AC P44827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN HI0694.
GN HI0694.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA Fleischmann R.D., Buit C.J., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.H.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC STRONG, TO E. COLI YMFEC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32752; AAC22354.1; -
CC TIGR; HI0694; -
DR PROSITE; PS01149; PSI_RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 240 AA; 27464 MW; 0AD99EC61F52C01C CRC64;

Query Match 6.8%; Score 15; DB 1; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 VYAAGRLDRDSEGLL 88
|||||
DB 84 VYAAGRLDRDSEGLL 98
|||||

RESULT 3
Y612_SYNY3
ID Y612_SYNY3 STANDARD; PRT; 261 AA.
AC P72581; -

```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 21.0 KD PROTEIN SLR0612.
GN SLR0612.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Bairoch A.;
RL Unpublished observations (SEP-1998).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 20 TO PRODUCE THIS ORF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90899; BAA16580.1; ALT_FRAME.
DR PROSITE; PS01149; PSI_RSU; 1.
KW Hypothetical protein.
SQ SEQUENCE 261 AA; 29561 MW; 601A453085C04A69 CRC64;

Query Match 5.4%; Score 12; DB 1; Length 261;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 EGRNRQVRMTA 187
| | | | | | | | | |
DB 175 EGRNRQVRMTA 186

RESULT 4
RLUB_BACSU STANDARD; PRT; 229 AA.
ID RLUB_BACSU STANDARD; PRT; 229 AA.
AC P35159.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70)
DE (PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLYASE).
GN RLUB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-168 / MABURG;
RA MEDLINE; 95020538.
RA Sorokin A.V., Zumbstein E., Azevedo V., Ehrlich S.D., Serror P.;
RT "The organization of the Bacillus subtilis 168 chromosome region
RT between the spoVA and serA genetic loci, based on sequence data.";
RL Mol. Microbiol. 10:385-395(1993).
RN [2]
RP CHARACTERIZATION.

RX MEDLINE; 98325071.
RA Niu L., Ofengand J.;
RL Unpublished results, cited by:
RL Conrad J., Sun D., Englund N., Ofengand J.;
RL J. Biol. Chem. 273:18562-18566(1998).
CC -!- FUNCTION: RESPONSIBLE FOR SYNTHESIS OF PSEUDOURIDINE FROM URACIL-
CC 2650 IN 23S RIBOSOMAL RNA.
CC -!- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE -> PSEUDOURIDINE
CC 5'-PHOSPHATE + H(2)O.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L09228; AAA67493.1; -
DR EMBL; Z99116; CAB14248.1; -
DR PIR; S45555; S45555.
DR SUBTILIST; BG10530; RLUB.
DR PROSITE; PS01149; PSI_RSU; 1.
DR PFAM; PF00849; YABO; 1.
DR PFAM; PF01479; S4; 1.
KW Lyase.
SQ SEQUENCE 229 AA; 26025 MW; 0CD54C2EE7A5ADE8 CRC64;

Query Match 4.5%; Score 10; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 EGRNRQVRM 185
| | | | | | | | | |
DB 184 EGRNRQVRM 193

RESULT 5
CH60_RICTS STANDARD; PRT; 555 AA.
ID CH60_RICTS STANDARD; PRT; 555 AA.
AC P16625.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (MAJOR ANTIGEN 58)
DE (58 KD ANTIGEN).
GN MOPA OR GROEL OR STA58.
OS Rickettsia tsutsugamushi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Orientia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90216005.
RA Stover C.K., Marana D.P., Dasch G.A., Oaks E.V.;
RT "Molecular cloning and sequence analysis of the sta58 major antigen
RT gene of Rickettsia tsutsugamushi: sequence homology and antigenic
RT comparison of sta58 to the 60-kilodalton family of stress proteins.";
RL Infect. Immun. 58:1360-1368(1990).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/

```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31887; AAA26393.1; -.
DR PIR; B41492; B41492.
DR HSSP; P06139; 1GRU.
DR PRINTS; P00298; CHAPERONIN60.
DR PRINTS; P00304; TCOMPLEXCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR PFAM; PF00118; cpn60_TCP1; 1.
KW Chaperone; ATP-binding; Antigen.
SQ SEQUENCE 555 AA; 59729 MW; 16400249D3FC03B4 CRC64;

Query Match 3.6%; Score 8; DB 1; Length 555;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 LAKLRNGV 131
DB 372 LAKLRNGV 379
|||||||

RESULT 6
REPA_AGRU STANDARD; PRT; 250 AA.
AC P15394;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE REPLICATING PROTEIN.
GN REPA.
OS Agrobacterium tumefaciens.
OC Plasmid pPR.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 88257036.
RA Gallie D.R., Kado C.I.;
RT "Minimal region necessary for autonomous replication of pTAR.";
RL J. Bacteriol. 170:3170-3176(1988).
CC -!- FUNCTION: REQUIRED FOR REPLICATION. IT LIKELY REGULATES PTAR
CC COPY NUMBER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21299; AAD15307.1; -.
DR PIR; A43662; A43662.
KW Plasmid; DNA replication.
SQ SEQUENCE 250 AA; 27987 MW; CCED106534831979 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 250;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 QPKRTG 108
DB 59 QPKRTG 65
|||||||

RESULT 7
STCV_EMENI STANDARD; PRT; 387 AA.
ID STCV_EMENI
AC Q00727;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

15-JUL-1999 (Rel. 38, Last annotation update)
PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV
(EC 1.1.1.1.-).
GN STCV.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Plectomycetes; Eurotiales;
OC Trichocomaceae; Emericella.
[1]
RN SEQUENCE FROM N.A.
RX STRAIN=FGSC 26;
RX MEDLINE; 96202293.
RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA Keller N.P., Adams T.H., Leonard T.J.;
RT "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT cluster in Aspergillus nidulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
CC -!- FUNCTION: MAY BE INVOLVED IN THE DEHYDRATION OF 5'-
CC HYDROXYAVERANTIN TO FORM AVERUTIN.
CC -!- PATHWAY: STERIGMATOCYSTIN BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U34740; AAC49206.1; -.
DR OXidoreductase.
FT ACT_SITE 148 148 HYDROGEN-BOND DONOR (POTENTIAL).
SQ SEQUENCE 387 AA; 43548 MW; 1621588273B8588C CRC64;

Query Match 3.2%; Score 7; DB 1; Length 387;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 ASLAKLR 128
DB 132 ASLAKLR 138
|||||||

RESULT 8
NF-M_PIG STANDARD; PRT; 454 AA.
ID NF-M_PIG
AC P08552;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M)
DE (FRAGMENT).
DE NF-M.
GN NF-M.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Metastodactyla; Suina; Suidae; Sus.
[1]
RN SEQUENCE.
RC TISSUE=SPINAL CORD;
RX MEDLINE; 85076594.
RA Geisler N., Fischer S., Vandekerckhove J., Plessmann U., Weber K.;
RT "Hybrid character of a large neurofilament protein (NF-M):
RT intermediate filament type sequence followed by a long and acidic
RT carboxy-terminal extension.";
RL EMBO J. 3:2701-2706(1984).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NF-M IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
```

CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC
 CC PIR: A05075; QPFGM.
 CC DR PROSITE: PS00226; IF: 1.
 CC DR PFAM: PF00038; filament; 1.
 CC KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
 CC ACetylation; Phosphorylation; Glycoprotein.
 CC FT MOD_RES 1 1
 CC FT DOMAIN 1 104 HEAD.
 CC FT DOMAIN 105 412 ROD.
 CC FT DOMAIN 413 >454 TAIL.
 CC FT DOMAIN 105 136 COIL 1A.
 CC FT DOMAIN 137 149 LINKER 1.
 CC FT DOMAIN 150 248 COIL 1B.
 CC FT DOMAIN 249 265 LINKER 12.
 CC FT DOMAIN 266 287 COIL 2A.
 CC FT DOMAIN 288 291 COIL 2B.
 CC FT DOMAIN 292 411 COIL 2B.
 CC FT CARBOHYD 47 47
 CC FT CARBOHYD 432 432 O-LINKED (GLCNAC) (BY SIMILARITY).
 CC FT NON_TER 454 454 O-LINKED (GLCNAC) (BY SIMILARITY).
 CC SEQUENCE 454 AA; 51854 MW; C809282BEC903A6B CRC64;

Query Match 3.2%; Score 7; DB 1; Length 454;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 SYTIDSL 209
 DB 1 SYTIDSL 7

RESULT 9
 PDI_ASPIG STANDARD; PRT; 515 AA.
 AC Q12730;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) (EC 5.3.4.1).
 GN PDI OR PDII.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Plectomycetes; Eurotiales;
 OC Trichocomaceae; anamorphic Trichocomaceae; Aspergillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 120.49 / N400;
 RX MEDLINE; 97174110.
 RA Ngiam C., Jeenes D.J., Archer D.B.;
 RT "Isolation and characterisation of a gene encoding protein disulphide
 RT isomerase, pdia, from Aspergillus niger."
 RL Curr. Genet. 31:133-138(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 3;
 RX MEDLINE; 97174110.
 RA Malpricht S.;
 RT Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 CC DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
 CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
 CC
 CC -!- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN
 CC DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
 CC
 CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY).
 CC
 CC -!- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: X98797; CAA67332.1;
 CC DR EMBL: X89449; CAA61619.1;
 CC DR HSSP: P07237; LMEK.
 CC DR PRINTS: PR00421; THIOREDOXIN.
 CC DR PROSITE: PS00014; ER_TARGET; 1.
 CC DR PROSITE: PS00194; THIOREDOXIN; 2.
 CC DR PFAM: PF00085; thior; 3.
 CC KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 CC FT SIGNAL 1 20 POTENTIAL.
 CC FT CHAIN 21 515 PROTEIN DISULFIDE ISOMERASE.
 CC FT DISULFID 54 57 REDOX-ACTIVE (BY SIMILARITY).
 CC FT DISULFID 389 392 REDOX-ACTIVE (BY SIMILARITY).
 CC FT SITE 512 515 PREVENT SECRETION FROM ER (POTENTIAL).
 CC SEQUENCE 515 AA; 56291 MW; 2B0058B788400AD9 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 515;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 GFPTLRL 196
 DB 434 GFPTLRL 440

RESULT 10
 TCPG_YEAST STANDARD; PRT; 534 AA.
 AC P39077;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE T-COMPLEX PROTEIN 1, GAMMA SUBUNIT (TCP-1-GAMMA) (CCT-GAMMA).
 GN CCT3 OR TCP3 OR BIN2 OR YJL014W OR J1336.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94377502.
 RA Chen X., Sullivan D.S., Huffaker T.C.;
 RT "Two yeast genes with similarity to TCP-1 are required for
 RT microtubule and actin function in vivo."
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9111-9115(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA To Van D., Perea J., Jacq C.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
 CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
 CC ACTIN AND TUBULIN. IN YEAST MAY PLAY A ROLE IN MITOTIC SPINDLE
 CC FORMATION.
 CC
 CC -!- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KD THAT
 CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.
 CC
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC
 CC -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U09480; AAA21658.1;
 CC DR EMBL: Z49289; CAA89305.1;
 CC DR PIR: S48875; S48875.
 CC DR HSSP: P48424; IASX.


```
DR SGD; L0002270; CCT6.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00750; TCPL1_1; 1.
DR PROSITE; PS00751; TCPL1_2; 1.
DR PROSITE; PS00995; TCPL1_3; 1.
DR PFAM; PF00118; cna60_TCP1; 1.
KW Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 534 AA; 58814 MW; 5C5AFF3D67D9A3B6 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 534;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 LVLNDG 94
| | | | |
DB 55 LVLNDG 61

RESULT 11
TCP2_YEAST
ID TCP2_YEAST STANDARD; PRT; 546 AA.
AC P39079;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE T-COMPLEX PROTEIN 1, ZETA SUBUNIT (TCP-1-ZETA) (CCT-2ZETA).
GN CTG6 OR TCP6 OR TCP20 OR YDR188W OR YD9395.21.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94308102.
RA Li W.-Z., Lin P., Frydman J., Boal T.R., Cardillo T.S., Richard L.M.,
RA Tsch D., Lichtman M.A., Hartl F.-U., Sherman F., Segel G.B.;
RA "tcp20, a subunit of the eukaryotic TRIC chaperonin from humans and
RT yeast.";
RL J. Biol. Chem. 269:18616-18622(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Murphy L., Harris D.E., Barrell B.G., Rajandream M.A.;
RA Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
CC APP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
CC ACTIN AND TUBULIN. IN YEAST MAY PLAY A ROLE IN MITOTIC SPINDLE
CC FORMATION.
CC -1- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KD THAT
CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; L27698; AAA33140.1;
CC EMBL; Z46727; CAA86694.1;
CC PIR; S48086; S48086.
CC HSSP; P48424; IASX.
CC YEPD; 6690;

DR SGD; L0002271; CCT6.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00750; TCPL1_1; 1.
DR PROSITE; PS00751; TCPL1_2; 1.
DR PROSITE; PS00995; TCPL1_3; 1.
DR PFAM; PF00118; cna60_TCP1; 1.
KW Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 546 AA; 59923 MW; 7D2A2AB4526D3DB4 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 546;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 AGRSTLK 66
| | | | |
DB 536 AGRSTLK 542

RESULT 12
SP20_YEAST
ID SP20_YEAST STANDARD; PRT; 604 AA.
AC P50875;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTION FACTOR SPT20.
GN SPT20 OR ADA5 OR YOL148C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RA Roberts S.M., Winston F.;
RA Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BWGI-7A;
RA Marcus G.A., Silverman N., Guarente L.;
RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Arino J., Casamayor A., Gamo F.J., Gancedo C., Lafuente M.J.,
RA Aldea M., Casas C., Herrero E.;
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 50-604 FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE; 96132030.
RA Casamayor A., Aldea M., Casas C., Herrero E., Gamo F.J.,
RA Lafuente M.J., Gancedo C., Arino J.;
RA "DNA sequence analysis of a 13 kbp fragment of the left arm of yeast
RT chromosome XV containing seven new open reading frames.";
RL Yeast 11:1281-1288(1995).
CC -1- FUNCTION: HFII/ADA1 AND SPT20/ADA5 MAY RECRUIT TATA BINDING
CC PROTEIN (TBP) AND POSSIBLY OTHER BASAL FACTORS TO BIND TO THE TATA
CC BOX WHILE ADA2, ADA3 AND GCN5 FUNCTION TO ACETYLATE NUCLEOSOMES
CC OPENING UP THE PROMOTER REGION.
CC -1- SUBUNIT: PART OF THE ADA/GCN5 COMPLEX THAT CONSISTS OF HFII/ADA1,
CC ADA2, ADA3, SPT20/ADA5 AND GCN5.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U22063; AAB07900.1;
CC EMBL; U43153; AAB07899.1;
CC EMBL; Z74890; CAA99169.1;
```



```
DR EMBL; 248239; CAA88279.1;
DR SGD; L0002593; SPT20.
KW Transcription regulation.
FT DOMAIN 157 162 POLY-GLN.
FT DOMAIN 235 240 POLY-SER.
FT DOMAIN 422 425 POLY-SER.
FT DOMAIN 454 463 POLY-ALA.
FT DOMAIN 552 559 POLY-ASN.
FT CONFLICT 293 Y -> S (IN REF. 4).
SQ SEQUENCE 604 AA; 67796 MW; 3D67937B65F9AA6A CRC64;

Query Match 3.2%; Score 7; DB 1; Length 604;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 DVTPEKN 222
DB 223 DVTPEKN 229

RESULT 13
COG9_RABIT STANDARD; PRT; 707 AA.
ID P41246;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 92 KD TYPE IV COLLAGENASE PRECURSOR (EC 3.4.24.35) (92 KD GELATINASE)
DE (MATRIX METALLOPROTEINASE-9) (MMP-9) (GELATINASE B).
GN MMP9.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-JAPANESE WHITE; TISSUE=BONE;
RX MEDLINE; 94253056.
RA Tezuka K.I., Nemoto K., Tezuka Y., Sato T., Ikeda Y., Kobori M.,
RA Kawashima H., Eguchi H., Hakeda Y., Kumegawa M.;
RT "Identification of matrix metalloproteinase 9 in rabbit osteoclasts.";
RL J. Biol. Chem. 269:15006-15009(1994).
RN [2]
RP SEQUENCE OF 1-171 FROM N.A.
RC STRAIN-NEW ZEALAND WHITE; TISSUE=LIVER;
RX MEDLINE; 95050662.
RA Fini M.E., Bartlett J.D., Matsubara M., Rinehart W.B., Mody M.K.,
RA Girard M.T., Rainville M.;
RT "The rabbit gene for 92-kDa matrix metalloproteinase. Role of API and
RT AP2 in cell type-specific transcription.";
RL J. Biol. Chem. 269:28620-28628(1994).
CC -1- FUNCTION: COULD PLAY A ROLE IN BONE OSTEOCLASTIC RESORPTION.
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF GELATIN TYPES I AND V AND COLLAGEN
CC TYPES IV AND V.
CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC -1- TISSUE SPECIFICITY: OSTEOCLASTS.
CC -1- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II
CC MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN
CC BINDS GELATIN.
CC -1- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE II-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC METALLOPROTEINASE) ALSO KNOWN AS MATRININ SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; D26514; BAA05520.1;
DR
```

```
DR EMBL; L36050; AAA64358.1;
DR HSSP; P08254; ISLN.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00138; MATRXIN.
DR PROSITE; PS00023; FIBRONECTIN.2; 3.
DR PROSITE; PS00024; HEMOPEXIN.1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR PFAM; PF00040; fn2; 3.
DR PFAM; PF00045; hemopexin; 4.
DR PFAM; PF00413; Peptidase_M10; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
KW Collagen degradation; Extracellular matrix; Signal.
FT SIGNAL 1 19
FT PROPEP 20 106
FT CHAIN 107 707
FT DOMAIN 97 104
FT DOMAIN 223 280
FT DOMAIN 281 339
FT DOMAIN 340 397
FT DOMAIN 511 707
FT METAL 401 401
FT ACT_SITE 402 402
FT METAL 405 405
FT METAL 411 411
FT CARBOHYD 88 88
FT CARBOHYD 120 120
FT CARBOHYD 127 127
FT DISULFID 516 704
FT CONFLICT 76 76
FT CONFLICT 100 102
FT SEQUENCE 707 AA; 78307 MW; 053BCE8DC4D4758F CRC64;

Query Match 3.2%; Score 7; DB 1; Length 707;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 TPEPQPT 41
DB 456 TPEPQPT 462

RESULT 14
ACOC_CUCMC STANDARD; PRT; 764 AA.
ID ACOC_CUCMC
AC Q42669;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ACONITASE (EC 4.2.1.3) (ACONITATE HYDRATASE) (CITRATE HYDRO-LYASE)
DE (FRAGMENT).
GN ACOC.
OS Cucumis melo var. conomon (Oriental pickling melon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Cucurbitales; Cucurbitaceae;
OC Cucumis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. CANTALOUPE; TISSUE=FRUIT;
RX MEDLINE; 95229629.
RA Peyret P., Perez P., Alric M.;
RT "Structure, genomic organization, and expression of the Arabidopsis
RT thaliana aconitase gene. Plant aconitase show significant homology
RT with mammalian iron-responsive element-binding protein.";
RL J. Biol. Chem. 270:8131-8137(1995).
CC -1- CATALYTIC ACTIVITY: CITRATE - CIS-ACONITATE + H(2)O.
CC -1- PATHWAY: GLYOXYLATE BYPASS, PLAYS A ROLE IN GLUCONEOGENESIS FROM
CC STORED OIL.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- MISCELLANEOUS: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE
CC (3FE-4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 11:25:05 ; Search time 16.42 Seconds
(without alignments)
937.404 Million cell updates/sec

Title: US-09-252-691-7056

Perfect score: 222
Sequence: 1 ALMROLIPNTMTTSFRK.....SYTLDLANGEMRWVTPKEN 222

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 225878 seqs, 59334122 residues

Word size : 0

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL12.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vtebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	3.6	341	5 P90743	P90743 caenorhabdi
2	8	3.6	448	2 O83284	O83284 treponema p
3	8	3.6	1494	5 O20943	O20943 caenorhabdi
4	8	3.6	1525	5 Q94137	Q94137 caenorhabdi
5	7	3.2	66	9 Q38044	Q38044 bacterioph
6	7	3.2	288	2 O51543	O51543 pseudomonas
7	7	3.2	312	2 P72806	P72806 synechocyst
8	7	3.2	352	2 O05938	O05938 pseudomonas
9	7	3.2	356	5 Q9Y179	Q9Y179 trichomonas
10	7	3.2	406	2 Q9ZMQ3	Q9ZMQ3 helicobacte
11	7	3.2	416	3 O00206	O00206 aspergillus
12	7	3.2	444	2 O86823	O86823 streptomyce
13	7	3.2	514	2 O25572	O25572 helicobacte
14	7	3.2	514	2 O9ZKT5	O9ZKT5 helicobacte
15	7	3.2	710	3 O14407	O14407 neurospora
16	7	3.2	766	10 Q39177	Q39177 arabidopsis
17	7	3.2	845	11 O63370	O63370 rattus norv
18	7	3.2	988	2 Q9X537	Q9X537 corynebacte
19	7	3.2	990	2 O06512	O06512 enterobacte

093787	rhizopus ol	3	093787	1112	3	093787
Q26372	tribolium c	5	Q26372	35	5	Q26372
Q78063	human immun	6	Q78063	35	6	Q78063
Q84304	human papil	12	Q84304	44	12	Q84304
Q84305	human papil	44	Q84305	44	44	Q84305
Q84309	human papil	44	Q84309	44	44	Q84309
Q84315	human papil	44	Q84315	44	44	Q84315
Q84914	human papil	44	Q84914	44	44	Q84914
Q97951	human immun	53	Q97951	53	53	Q97951
Q29179	archaeoglob	1	Q29179	69	1	Q29179
Q35310	oryza sativ	8	Q35310	71	8	Q35310
Q04252	human papil	12	Q04252	79	12	Q04252
Q43687	homo sapien	4	Q43687	81	4	Q43687
Q80286	human immun	12	Q80286	92	12	Q80286
Q92b2	streptomyce	2	Q92b2	93	2	Q92b2
Q47019	escherichia	2	Q47019	94	2	Q47019
Q9wu87	mus musculu	11	Q9wu87	112	11	Q9wu87
Q33159	mycobacteri	2	Q33159	113	2	Q33159
Q80978	arabidopsis	10	Q80978	117	10	Q80978
Q56628	helicoverpa	12	Q56628	118	12	Q56628
Q56631	tenebrio mo	12	Q56631	118	12	Q56631
Q56632	apis cerana	12	Q56632	118	12	Q56632
Q56635	anticarsia	12	Q56635	118	12	Q56635
Q01694	caenorhabdl	5	Q01694	119	5	Q01694
Q81850	human papil	12	Q81850	128	12	Q81850
Q77762	canis famil	6	Q77762	132	6	Q77762
Q81852	human papil	12	Q81852	132	12	Q81852
Q9xs58	tursiops tr	6	Q9xs58	133	6	Q9xs58
P93856	skeletonema	10	P93856	137	10	P93856
Q19840	caenorhabdl	5	Q19840	138	5	Q19840
Q53856	spiroplasma	2	Q53856	139	2	Q53856
Q50881	myxococcus	2	Q50881	140	2	Q50881
Q08687	saccharomyc	3	Q08687	141	3	Q08687
Q00917	plasmodium	5	Q00917	142	5	Q00917
Q00921	plasmodium	5	Q00921	142	5	Q00921
Q9x864	streptomyce	2	Q9x864	145	2	Q9x864
O52243	acinetobact	2	O52243	146	2	O52243
O52245	acinetobact	2	O52245	146	2	O52245
Q43899	acinetobact	2	Q43899	146	2	Q43899
Q07880	saccharomyc	3	Q07880	148	3	Q07880
Q26170	plasmodium	5	Q26170	151	5	Q26170
Q9xc15	shigella fl	2	Q9xc15	153	2	Q9xc15
O21999	bacterioph	9	O21999	154	9	O21999
Q05745	saccharomyc	3	Q05745	156	3	Q05745
O55052	streptococc	2	O55052	171	2	O55052
Q03839	homo sapien	4	Q03839	171	4	Q03839
O00923	plasmodium	5	O00923	171	5	O00923
O00918	plasmodium	5	O00918	172	5	O00918
O00919	plasmodium	5	O00919	172	5	O00919
O00920	plasmodium	5	O00920	172	5	O00920
O00922	plasmodium	5	O00922	172	5	O00922
O26185	plasmodium	5	O26185	172	5	O26185
O34479	bacillus su	2	O34479	173	2	O34479
P73586	synechocyst	2	P73586	173	2	P73586
O64175	bacterioph	9	O64175	173	9	O64175
Q9x604	salmonella	2	Q9x604	180	2	Q9x604
O74882	schizosacch	3	O74882	180	3	O74882
Q84877	human polio	12	Q84877	182	12	Q84877
O66304	coxsackievi	12	O66304	183	12	O66304
O66305	coxsackievi	12	O66305	183	12	O66305
O66306	coxsackievi	12	O66306	183	12	O66306
O66307	coxsackievi	12	O66307	183	12	O66307
O66308	coxsackievi	12	O66308	183	12	O66308
O66309	coxsackievi	12	O66309	183	12	O66309
O66310	coxsackievi	12	O66310	183	12	O66310
O66311	coxsackievi	12	O66311	183	12	O66311
O66312	coxsackievi	12	O66312	183	12	O66312
O66313	coxsackievi	12	O66313	183	12	O66313
O66314	coxsackievi	12	O66314	183	12	O66314
O66315	coxsackievi	12	O66315	183	12	O66315
O66316	coxsackievi	12	O66316	183	12	O66316
O66317	coxsackievi	12	O66317	183	12	O66317
O66318	coxsackievi	12	O66318	183	12	O66318

93	6	2.7	183	12	066319	066319 coxsackievi	166	6	2.7	315	5	000799	000799 plasmodium
94	6	2.7	183	12	066320	066320 coxsackievi	167	6	2.7	315	5	000802	000802 plasmodium
95	6	2.7	183	12	090207	090207 coxsackievi	168	6	2.7	315	5	000803	000803 plasmodium
96	6	2.7	183	12	090208	090208 coxsackievi	169	6	2.7	317	2	P96034	P96034 synechococ
97	6	2.7	183	12	090209	090209 coxsackievi	170	6	2.7	319	2	034966	034966 bacillus su
98	6	2.7	189	10	092006	092006 pseudotsuga	171	6	2.7	319	2	059553	059553 mycobacteri
99	6	2.7	190	2	086520	086520 streptomyce	172	6	2.7	321	2	P96240	P96240 mycobacteri
100	6	2.7	190	10	P93855	P93855 skeletonema	173	6	2.7	322	4	075370	075370 homo sapien
101	6	2.7	194	5	016565	016565 caenorhabdi	174	6	2.7	324	5	09XXP7	09XXP7 caenorhabdi
102	6	2.7	199	12	084301	084301 human papil	175	6	2.7	328	2	068786	068786 yersinia pe
103	6	2.7	209	2	054128	054128 streptomyce	176	6	2.7	328	5	045299	045299 caenorhabdi
104	6	2.7	213	2	09X6F7	09X6F7 campylobact	177	6	2.7	329	5	P90613	P90613 toxoplasma
105	6	2.7	213	2	09X6F5	09X6F5 campylobact	178	6	2.7	331	5	017810	017810 caenorhabdi
106	6	2.7	216	3	Q12143	Q12143 saccharomyc	179	6	2.7	333	5	000795	000795 plasmodium
107	6	2.7	221	2	P71487	P71487 methyloicr	180	6	2.7	336	2	069207	069207 actinosynne
108	6	2.7	223	12	09YMN7	09YMN7 lymantria d	181	6	2.7	338	3	043083	043083 schizosacch
109	6	2.7	225	10	042436	042436 brassica ra	182	6	2.7	338	5	000791	000791 plasmodium
110	6	2.7	228	2	050447	050447 mycobacteri	183	6	2.7	338	5	000793	000793 plasmodium
111	6	2.7	228	2	092506	092506 streptomyce	184	6	2.7	338	5	096270	096270 plasmodium
112	6	2.7	233	2	09X0E1	09X0E1 thermotoga	185	6	2.7	339	5	000792	000792 plasmodium
113	6	2.7	235	2	051240	051240 borrelia bu	186	6	2.7	339	5	000800	000800 plasmodium
114	6	2.7	235	2	09XC14	09XC14 shigella fl	187	6	2.7	342	2	083088	083088 treponema p
115	6	2.7	238	2	P74198	P74198 synechocyst	188	6	2.7	344	5	045427	045427 caenorhabdi
116	6	2.7	240	5	09XXH8	09XXH8 thermotoga	189	6	2.7	345	2	086241	086241 haemophilus
117	6	2.7	240	5	09XXH3	09XXH3 babesia big	190	6	2.7	345	2	005599	005599 pseudomonas
118	6	2.7	241	2	086961	086961 thermotoga	191	6	2.7	351	10	P93547	P93547 spinacia ol
119	6	2.7	242	2	P96892	P96892 mycobacteri	192	6	2.7	353	12	085447	085447 rice dwarf
120	6	2.7	245	12	075832	075832 human immun	193	6	2.7	355	12	09YUG8	09YUG8 rhesus cyto
121	6	2.7	246	12	075834	075834 human immun	194	6	2.7	356	5	09XTB2	09XTB2 caenorhabdi
122	6	2.7	250	2	034363	034363 streptomyce	195	6	2.7	360	2	053411	053411 mycobacteri
123	6	2.7	251	2	048792	048792 listeria mo	196	6	2.7	361	2	087547	087547 treponema d
124	6	2.7	253	1	09XAV7	09XAV7 aeropyrum p	197	6	2.7	363	2	044198	044198 agrobacteri
125	6	2.7	254	2	030757	030757 borrelia bu	198	6	2.7	364	12	081549	081549 hepatitis c
126	6	2.7	256	4	035466	035466 homo sapien	199	6	2.7	364	12	081551	081551 hepatitis c
127	6	2.7	257	2	052617	052617 pseudomonas	200	6	2.7	365	12	081716	081716 hepatitis c
128	6	2.7	257	2	059009	059009 pseudomonas	201	6	2.7	367	12	072468	072468 nelson bay
129	6	2.7	257	2	095648	095648 pseudomonas	202	6	2.7	368	2	053047	053047 lactobacill
130	6	2.7	257	12	P89923	P89923 bacterioph	203	6	2.7	370	4	09Y360	09Y360 homo sapien
131	6	2.7	258	2	P71968	P71968 mycobacteri	204	6	2.7	371	2	0927K1	0927K1 chlamydia p
132	6	2.7	258	2	P96163	P96163 vibrio furn	205	6	2.7	372	12	P89439	P89439 herpes simp
133	6	2.7	261	2	054172	054172 streptomyce	206	6	2.7	377	1	059445	059445 pyrococcus
134	6	2.7	262	12	086975	086975 phocid herp	207	6	2.7	377	10	048680	048680 arabidopsis
135	6	2.7	265	2	066158	066158 streptomyce	208	6	2.7	382	6	09X561	09X561 bos taurus
136	6	2.7	266	2	051400	051400 pseudomonas	209	6	2.7	382	1	09YFP0	09YFP0 aeropyrum p
137	6	2.7	268	2	005684	005684 mycobacteri	210	6	2.7	383	2	09WZ93	09WZ93 thermotoga
138	6	2.7	269	2	084991	084991 rhodococcus	211	6	2.7	386	2	09X829	09X829 streptomyce
139	6	2.7	269	2	09X9X6	09X9X6 streptomyce	212	6	2.7	386	13	091839	091839 xenopus lae
140	6	2.7	270	5	093721	093721 caenorhabdi	213	6	2.7	390	1	09Y809	09Y809 aeropyrum p
141	6	2.7	271	2	09X213	09X213 thermotoga	214	6	2.7	390	2	09Z1D8	09Z1D8 actinobacil
142	6	2.7	273	2	049993	049993 mycobacteri	215	6	2.7	390	13	09W6U5	09W6U5 fugu rubrip
143	6	2.7	274	2	P71671	P71671 mycobacteri	216	6	2.7	391	2	052822	052822 amycolatops
144	6	2.7	276	10	065599	065599 arabidopsis	217	6	2.7	391	2	052822	052822 amycolatops
145	6	2.7	277	5	027045	027045 theileria p	218	6	2.7	391	5	017327	017327 caenorhabdi
146	6	2.7	281	1	059568	059568 pyrococcus	219	6	2.7	393	1	09YBB6	09YBB6 aeropyrum p
147	6	2.7	281	2	087598	087598 streptomyce	220	6	2.7	402	5	003486	003486 physarum po
148	6	2.7	281	2	09WZB5	09WZB5 thermotoga	221	6	2.7	402	13	09W5Z9	09W5Z9 fugu rubrip
149	6	2.7	285	10	09ZWP5	09ZWP5 robinia pse	222	6	2.7	403	2	045380	045380 bordetella
150	6	2.7	288	1	058132	058132 pyrococcus	223	6	2.7	403	2	069941	069941 streptomyce
151	6	2.7	294	2	09X7T0	09X7T0 streptomyce	224	6	2.7	406	2	052823	052823 amycolatops
152	6	2.7	294	13	09YH87	09YH87 oryzias lat	225	6	2.7	407	2	031557	031557 bacillus su
153	6	2.7	296	2	09X901	09X901 streptomyce	226	6	2.7	407	3	059721	059721 schizosacch
154	6	2.7	298	12	041026	041026 paramacium	227	6	2.7	407	10	082257	082257 arabidopsis
155	6	2.7	300	2	066788	066788 aquifex ae	228	6	2.7	408	2	034765	034765 bacillus su
156	6	2.7	300	9	09ZXD7	09ZXD7 bacterioph	229	6	2.7	409	5	09XZF6	09XZF6 chaetoteru
157	6	2.7	301	5	018331	018331 hemiceintrot	230	6	2.7	412	12	065354	065354 autographa
158	6	2.7	304	5	008001	008001 plasmodium	231	6	2.7	413	5	062419	062419 caenorhabdi
159	6	2.7	307	2	09ZBV0	09ZBV0 streptomyce	232	6	2.7	414	11	061395	061395 mus musculu
160	6	2.7	312	1	037397	027397 methanobact	233	6	2.7	414	11	005732	005732 mus musculu
161	6	2.7	312	5	000796	000796 plasmodium	234	6	2.7	416	2	025667	025667 helicobacte
162	6	2.7	312	5	000797	000797 plasmodium	235	6	2.7	416	2	092M22	092M22 helicobacte
163	6	2.7	313	2	088034	088034 streptomyce	236	6	2.7	419	2	087999	087999 bordetella
164	6	2.7	315	5	000794	000794 plasmodium	237	6	2.7	421	11	P97610	P97610 rattus norv
165	6	2.7	315	5	000798	000798 plasmodium	238	6	2.7	422	2	054505	054505 salmonella

239	6	2.7	425	2	Q9WYM6	Q9WYM6 thermotoga	312	6	2.7	547	1	Q29248	O29248 archaeoglob
240	6	2.7	433	5	Q45500	O45500 caenorhabdi	313	6	2.7	552	5	Q45812	O45812 caenorhabdi
241	6	2.7	436	1	Q9Y964	Q9Y964 aeropyrum p	314	6	2.7	553	5	Q26642	Q26642 sarcophaga
242	6	2.7	436	10	Q9XG60	Q9XG60 selaginella	315	6	2.7	554	12	Q56285	O56285 human herpe
243	6	2.7	436	10	Q9XGF9	Q9XGF9 selaginella	316	6	2.7	561	2	Q34105	O34105 salmonella
244	6	2.7	437	11	Q35249	Q35249 rattus norv	317	6	2.7	563	2	Q30916	O30916 salmonella
245	6	2.7	437	10	Q81783	O81783 arabisdopsis	318	6	2.7	564	5	Q27937	Q27937 loligo forb
246	6	2.7	438	11	Q9WV78	Q9WV78 rattus norv	319	6	2.7	566	3	P87142	P87142 schizosacch
247	6	2.7	439	4	Q44314	Q44314 homo sapien	320	6	2.7	569	1	Q28188	Q28188 archaeoglob
248	6	2.7	439	5	Q9Y122	Q9Y122 drosophila	321	6	2.7	573	5	O17654	O17654 caenorhabdi
249	6	2.7	444	1	Q27792	Q27792 methanobact	322	6	2.7	584	13	Q98921	Q98921 gallus gall
250	6	2.7	445	2	Q55937	Q55937 synechocyst	323	6	2.7	584	13	Q90989	Q90989 gallus gall
251	6	2.7	447	2	P76097	P76097 escherichia	324	6	2.7	587	2	O53417	O53417 mycobacteri
252	6	2.7	447	11	Q62162	Q62162 mus musculu	325	6	2.7	587	10	Q23340	Q23340 arabisdopsis
253	6	2.7	448	2	Q924F7	Q924F7 escherichia	326	6	2.7	591	4	Q01720	Q01720 homo sapien
254	6	2.7	450	2	O50924	O50924 borrelia bu	327	6	2.7	596	5	O62306	O62306 caenorhabdi
255	6	2.7	456	5	O17120	O17120 caenorhabdi	328	6	2.7	600	2	Q92DQ1	Q92DQ1 rickettsia
256	6	2.7	461	1	Q9YB22	Q9YB22 aeropyrum p	329	6	2.7	611	5	P90858	P90858 caenorhabdi
257	6	2.7	465	2	Q9XBP5	Q9XBP5 myxococcus	330	6	2.7	614	3	Q01203	Q01203 melampsora
258	6	2.7	466	2	Q56291	Q56291 thioacillu	331	6	2.7	617	5	O96437	O96437 elmeria ten
259	6	2.7	467	11	O70527	O70527 cavia porce	332	6	2.7	622	5	O00926	O00926 trypanosoma
260	6	2.7	469	5	O77466	O77466 mycobacteri	333	6	2.7	622	5	O17082	O17082 caenorhabdi
261	6	2.7	474	2	O53446	O53446 mycobacteri	334	6	2.7	626	13	Q98922	Q98922 gallus gall
262	6	2.7	474	2	Q9X789	Q9X789 mycobacteri	335	6	2.7	626	13	Q90880	Q90880 gallus gall
263	6	2.7	485	5	Q9XW17	Q9XW17 caenorhabdi	336	6	2.7	636	5	Q26182	Q26182 plasmodium
264	6	2.7	490	2	O83778	O83778 treponema p	337	6	2.7	638	2	O67583	O67583 aquifex aeo
265	6	2.7	490	2	Q9XAL7	Q9XAL7 streptomyc	338	6	2.7	640	3	Q07528	Q07528 saccharomyc
266	6	2.7	497	5	O76191	O76191 dirofilaria	339	6	2.7	643	12	Q84876	Q84876 human polio
267	6	2.7	498	2	O54116	O54116 streptomyc	340	6	2.7	647	2	O52071	O52071 lactobacill
268	6	2.7	500	12	Q9WB51	Q9WB51 human papil	341	6	2.7	648	12	Q96714	Q96714 chlorella v
269	6	2.7	500	12	Q9WB50	Q9WB50 human papil	342	6	2.7	649	12	O41043	O41043 parametium
270	6	2.7	500	12	Q9WB89	Q9WB89 human papil	343	6	2.7	651	5	Q27150	Q27150 oxytricha n
271	6	2.7	500	12	Q9WB88	Q9WB88 human papil	344	6	2.7	663	1	Q28469	Q28469 archaeoglob
272	6	2.7	500	12	Q9WB87	Q9WB87 human papil	345	6	2.7	672	2	Q45018	Q45018 borrelia bu
273	6	2.7	500	12	Q9WB86	Q9WB86 human papil	346	6	2.7	672	2	Q45041	Q45041 borrelia bu
274	6	2.7	500	12	Q9WB85	Q9WB85 human papil	347	6	2.7	674	5	P90755	P90755 caenorhabdi
275	6	2.7	500	12	Q9W9C6	Q9W9C6 human papil	348	6	2.7	679	2	O83759	O83759 treponema p
276	6	2.7	501	2	O50157	O50157 streptococc	349	6	2.7	683	5	O00838	O00838 leishmania
277	6	2.7	501	12	Q37391	Q37391 common chim	350	6	2.7	684	2	Q35846	Q35846 synechocyst
278	6	2.7	503	12	Q82010	Q82010 human papil	351	6	2.7	687	5	O62527	O62527 drosophila
279	6	2.7	503	12	Q81958	Q81958 human papil	352	6	2.7	692	5	O45101	O45101 caenorhabdi
280	6	2.7	503	12	Q9WNN4	Q9WNN4 human papil	353	6	2.7	692	5	O45016	O45016 borrelia bu
281	6	2.7	504	12	O81961	O81961 human papil	354	6	2.7	693	2	Q44937	Q44937 borrelia bu
282	6	2.7	504	13	Q98923	Q98923 gallus gall	355	6	2.7	693	2	Q45043	Q45043 borrelia bu
283	6	2.7	505	12	Q9YXD1	Q9YXD1 human papil	356	6	2.7	695	2	O45163	O45163 borrelia ga
284	6	2.7	506	12	Q9XEG4	Q9XEG4 hordeum vul	357	6	2.7	695	2	O52753	O52753 ruminococcu
285	6	2.7	506	12	O82004	O82004 human papil	358	6	2.7	698	1	O26996	O26996 methanobact
286	6	2.7	507	10	P93182	P93182 hordeum vul	359	6	2.7	698	5	O44850	O44850 caenorhabdi
287	6	2.7	507	10	Q9XEG5	Q9XEG5 hordeum vul	360	6	2.7	707	5	O00839	O00839 leishmania
288	6	2.7	508	12	Q81971	Q81971 human papil	361	6	2.7	714	12	Q83072	Q83072 leishmania
289	6	2.7	509	10	P93183	P93183 hordeum vul	362	6	2.7	717	2	Q48553	Q48553 leuconostoc
290	6	2.7	510	11	Q921K8	Q921K8 mus musculu	363	6	2.7	717	4	Q9Y4W2	Q9Y4W2 homo sapien
291	6	2.7	510	12	P89427	P89427 human papil	364	6	2.7	728	3	O94439	O94439 schizosacch
292	6	2.7	511	2	Q924W7	Q924W7 streptomyc	365	6	2.7	730	11	O70585	O70585 mus musculu
293	6	2.7	511	10	O81709	O81709 arabisdopsis	366	6	2.7	732	5	Q24557	Q24557 drosophila
294	6	2.7	512	2	P94684	P94684 chlamydia t	367	6	2.7	738	2	Q59490	Q59490 lactobacill
295	6	2.7	512	2	Q53769	Q53769 streptomyc	368	6	2.7	738	4	O60327	O60327 homo sapien
296	6	2.7	512	5	O01870	O01870 caenorhabdi	369	6	2.7	740	3	O93967	O93967 amanita mus
297	6	2.7	514	1	Q9YGG4	Q9YGG4 aeropyrum p	370	6	2.7	746	11	O88408	O88408 mus musculu
298	6	2.7	514	2	Q9XAJ9	Q9XAJ9 streptomyc	371	6	2.7	752	12	Q84875	Q84875 human polio
299	6	2.7	514	5	P91036	P91036 caenorhabdi	372	6	2.7	756	2	Q9X3F3	Q9X3F3 pseudomonas
300	6	2.7	517	5	P91137	P91137 caenorhabdi	373	6	2.7	757	12	Q85082	Q85082 human polio
301	6	2.7	524	10	Q23240	Q23240 arabisdopsis	374	6	2.7	760	10	O48795	O48795 arabisdopsis
302	6	2.7	525	4	Q9Y2K4	Q9Y2K4 homo sapien	375	6	2.7	766	2	O33466	O33466 pseudomonas
303	6	2.7	526	2	Q928Y6	Q928Y6 chlamydia p	376	6	2.7	766	2	O33466	O33466 pseudomonas
304	6	2.7	529	12	O90730	O90730 human papil	377	6	2.7	770	2	O85783	O85783 myxococcus
305	6	2.7	532	3	Q92205	Q92205 botrytis ci	378	6	2.7	773	10	Q96325	Q96325 arabisdopsis
306	6	2.7	533	2	Q53903	Q53903 streptomyc	379	6	2.7	775	10	O64990	O64990 arabisdopsis
307	6	2.7	534	12	Q82003	Q82003 human papil	380	6	2.7	775	10	P92941	P92941 arabisdopsis
308	6	2.7	536	4	Q9Y4Y5	Q9Y4Y5 homo sapien	381	6	2.7	776	12	Q98167	Q98167 lapine tota
309	6	2.7	537	5	Q25332	Q25332 leishmania	382	6	2.7	776	12	Q98168	Q98168 lapine tota
310	6	2.7	540	5	Q24281	Q24281 drosophila	383	6	2.7	780	10	P92942	P92942 arabisdopsis
311	6	2.7	541	10	Q40283	Q40283 manihot esc	384	6	2.7	783	10	O65383	O65383 arabisdopsis

385	6	2.7	786	2	032044	032044 bacillus su	458	6	2.7	1494	5	Q94885	Q94885 drosophila
386	6	2.7	786	10	Q9XF71	Q9xf71 nicotiana t	459	6	2.7	1510	4	Q9Y4G3	Q9Y4g3 homo sapien
387	6	2.7	790	5	001264	001264 caenorhabdi	460	6	2.7	1516	4	P78354	P78354 homo sapien
388	6	2.7	802	5	019764	019764 caenorhabdi	461	6	2.7	1524	5	O15830	O15830 trypanosoma
389	6	2.7	805	3	094133	094133 piromyces s	462	6	2.7	1532	4	P78544	P78544 homo sapien
390	6	2.7	814	3	Q05958	Q05958 saccharomyc	463	6	2.7	1549	4	O60706	O60706 homo sapien
391	6	2.7	825	2	Q48760	Q48760 listeria mo	464	6	2.7	1549	4	O60707	O60707 homo sapien
392	6	2.7	835	2	Q9X0A6	Q9x0a6 thermotoga	465	6	2.7	1562	2	Q92G13	Q92g13 streptomyce
393	6	2.7	840	4	O14587	O14587 homo sapien	466	6	2.7	1592	5	Q19386	Q19386 caenorhabdi
394	6	2.7	841	2	Q92388	Q92388 helicobacte	467	6	2.7	1684	4	O00310	O00310 homo sapien
395	6	2.7	842	2	Q25950	Q25950 helicobacte	468	6	2.7	1710	3	O74135	O74136 candida alb
396	6	2.7	884	10	Q42847	Q42847 hordium vul	469	6	2.7	1724	3	Q93927	Q93927 cryptococcu
397	6	2.7	873	2	Q05298	Q05298 bradyrhizob	470	6	2.7	1726	5	Q02569	Q02569 plasmodium
398	6	2.7	873	10	Q42739	Q42739 flaveria tr	471	6	2.7	1741	4	O00498	O00498 homo sapien
399	6	2.7	877	10	Q04297	Q04297 chlorella v	472	6	2.7	1751	5	Q26194	Q26194 plasmodium
400	6	2.7	879	3	Q07807	Q07807 saccharomyc	473	6	2.7	1767	5	Q24495	Q24495 drosophila
401	6	2.7	879	11	Q9WV91	Q9wv91 mus musculu	474	6	2.7	1781	4	Q99970	Q99970 homo sapien
402	6	2.7	884	10	Q23928	Q23928 eleocharis	475	6	2.7	1820	5	Q9XTF0	Q9xtf0 caenorhabdi
403	6	2.7	887	10	O82032	O82032 oryza sativ	476	6	2.7	1857	4	O95153	O95153 homo sapien
404	6	2.7	888	5	Q25336	Q25336 leishmania	477	6	2.7	1862	5	Q02090	Q02090 caenorhabdi
405	6	2.7	890	11	Q921A0	Q921a0 cavia porce	478	6	2.7	2051	5	O44328	O44328 hiruodo medi
406	6	2.7	896	11	O64146	Q64146 rattus norv	479	6	2.7	2080	4	O75923	O75923 homo sapien
407	6	2.7	907	10	Q42738	Q42738 flaveria tr	480	6	2.7	2091	3	P78616	P78616 emericella
408	6	2.7	912	12	Q90278	Q90278 chimpanzee	481	6	2.7	2118	5	O76904	O76904 drosophila
409	6	2.7	913	4	Q9Y2K6	Q9y2k6 homo sapien	482	6	2.7	2206	12	Q84792	Q84792 human polio
410	6	2.7	919	12	Q9WPP0	Q9wpp0 chimpanzee	483	6	2.7	2207	12	Q98595	Q98595 human polio
411	6	2.7	921	5	O42253	Q42253 drosophila	484	6	2.7	2221	12	O84865	O84865 human polio
412	6	2.7	947	10	Q23927	Q23927 eleocharis	485	6	2.7	2251	12	O11388	O11388 porcine rub
413	6	2.7	947	10	O41847	Q41847 oryza sativ	486	6	2.7	2342	12	O65980	O65980 cherry capi
414	6	2.7	947	10	O41847	Q41847 zea mays (m	487	6	2.7	2433	12	O91464	O91464 aichi virus
415	6	2.7	950	12	Q66678	Q66678 equine herp	488	6	2.7	2514	5	Q9Y061	Q9Y061 caenorhabdi
416	6	2.7	960	10	Q23404	Q23404 arabidopsis	489	6	2.7	2533	5	Q27183	Q27183 paramecium
417	6	2.7	980	12	Q66682	Q66682 equine herp	490	6	2.7	2533	5	P90589	P90589 paramecium
418	6	2.7	990	4	Q15206	Q15206 homo sapien	491	6	2.7	2619	5	Q22258	Q22258 caenorhabdi
419	6	2.7	990	10	Q40699	Q40699 oryza sativ	492	6	2.7	2628	3	Q06179	Q06179 saccharomyc
420	6	2.7	997	2	Q9Z4T1	Q9z4t1 bacillus sp	493	6	2.7	2705	13	Q9W6V6	Q9w6v6 gallus gall
421	6	2.7	1003	1	O58064	O58064 pyrococcus	494	6	2.7	2731	2	Q9Z3T9	Q9z3t9 pseudomonas
422	6	2.7	1003	5	Q9XXR9	Q9xxr9 caenorhabdi	495	6	2.7	2731	11	Q9WT84	Q9wts4 mus musculu
423	6	2.7	1010	13	Q9YHU7	Q9yhu7 xenopus lae	496	6	2.7	2809	5	O61230	O61230 lytechinus
424	6	2.7	1014	5	O61702	O61702 limulus pol	497	6	2.7	2894	1	O58791	O58791 methanococ
425	6	2.7	1038	2	O05924	O05924 neisseria m	498	6	2.7	3011	12	O36579	O36579 hepatitis c
426	6	2.7	1049	5	O45572	O45572 caenorhabdi	499	6	2.7	3011	12	O36608	O36608 hepatitis c
427	6	2.7	1067	13	Q9YHU6	Q9yhu6 xenopus lae	500	6	2.7	3011	12	O36609	O36609 hepatitis c
428	6	2.7	1079	2	Q95999	Q95999 n carbamoyl	501	6	2.7	3011	12	O36610	O36610 hepatitis c
429	6	2.7	1103	12	O41928	O41928 murine herp	502	6	2.7	3052	12	Q03463	Q03463 hepatitis c
430	6	2.7	1108	3	Q9Y7P5	Q9y7p5 schizosacch	503	6	2.7	3124	12	Q82933	Q82933 johnsongras
431	6	2.7	1139	2	O54073	O54073 synchococch	504	6	2.7	3161	2	O66237	O66237 citrus tris
432	6	2.7	1142	12	P89462	P89462 herpes simp	505	6	2.7	3161	2	O54511	O54511 versinia en
433	6	2.7	1194	13	Q9W737	Q9w737 gallus gall	506	6	2.7	3163	2	O9Z373	O9z373 versinia pe
434	6	2.7	1218	4	Q05331	Q05331 homo sapien	507	6	2.7	3413	2	O54593	O54593 amycolatops
435	6	2.7	1232	2	P94692	P94692 desulfovibr	508	6	2.7	3851	4	O43161	O43161 homo sapien
436	6	2.7	1234	5	O01505	O01505 caenorhabdi	509	6	2.7	4919	2	Q9ZHL0	Q9zh10 haemophilus
437	6	2.7	1240	13	P79773	P79773 gallus gall	510	6	2.7	6658	5	O76281	O76281 drosophila
438	6	2.7	1243	11	O35954	O35954 mus musculu	511	6	2.7	9	4	O95574	O95574 homo sapien
439	6	2.7	1244	4	O00562	O00562 homo sapien	512	5	2.3	15	4	P78533	P78533 homo sapien
440	6	2.7	1270	5	O19736	O19736 caenorhabdi	513	5	2.3	16	11	Q9Z0J2	Q9z0j2 mus musculu
441	6	2.7	1288	3	O13682	O13682 schizosacch	514	5	2.3	19	8	O63058	O63058 latraea cl
442	6	2.7	1329	10	Q9ZUK1	Q9zuk1 arabidopsis	515	5	2.3	21	12	Q82392	Q82392 human t-cel
443	6	2.7	1330	12	Q04534	Q04534 pseudorabie	516	5	2.3	21	12	Q82410	Q82410 human t-cel
444	6	2.7	1344	11	O60697	O60697 mus musculu	517	5	2.3	21	12	Q82411	Q82411 human t-cel
445	6	2.7	1350	5	Q27571	Q27571 drosophila	518	5	2.3	21	12	Q82412	Q82412 human t-cel
446	6	2.7	1364	4	O75981	O75981 homo sapien	519	5	2.3	26	10	Q40536	Q40536 nicotiana t
447	6	2.7	1381	11	P97846	P97846 rattus norv	520	5	2.3	32	11	Q64085	Q64085 mus musculu
448	6	2.7	1382	4	Q14152	Q14152 homo sapien	521	5	2.3	33	12	O73052	O73052 hepatitis c
449	6	2.7	1384	4	P78357	P78357 homo sapien	522	5	2.3	33	12	O73053	O73053 hepatitis c
450	6	2.7	1385	5	O26388	Q26388 drosophila	523	5	2.3	34	5	O61261	O61261 acyrtosiph
451	6	2.7	1385	11	O54991	O54991 mus musculu	524	5	2.3	34	12	O71809	O71809 human immun
452	6	2.7	1387	13	O57512	O57512 anas platyr	525	5	2.3	34	12	O71813	O71813 human immun
453	6	2.7	1389	5	Q24591	Q24591 drosophila	526	5	2.3	34	12	O71814	O71814 human immun
454	6	2.7	1389	13	Q90240	Q90240 anas sp. (d	527	5	2.3	34	12	O79463	O79463 human immun
455	6	2.7	1395	8	Q35059	Q35059 marchantia	528	5	2.3	34	12	O79493	O79493 human immun
456	6	2.7	1413	4	O75696	O75696 homo sapien	529	5	2.3	34	12	O91354	O91354 human immun
457	6	2.7	1475	10	Q9XEP3	Q9xep3 sorghum bic	530	5	2.3	34	12	O91355	O91355 human immun

531	34	12	O91356	human	immun	604	5	2.3	35	12	Q78145	human	immun
532	34	12	O91357	human	immun	605	5	2.3	35	12	Q78175	human	immun
533	34	12	O91358	human	immun	606	5	2.3	35	12	Q78186	human	immun
534	34	12	O91359	human	immun	607	5	2.3	35	12	Q78206	human	immun
535	34	12	O91360	human	immun	608	5	2.3	35	12	Q78213	human	immun
536	34	12	O91361	human	immun	609	5	2.3	35	12	Q78215	human	immun
537	34	12	O91365	human	immun	610	5	2.3	35	12	Q78215	human	immun
538	34	12	O91367	human	immun	611	5	2.3	35	12	Q79484	human	immun
539	34	12	O91368	human	immun	612	5	2.3	35	12	Q79500	human	immun
540	35	8	Q32706	human	immun	613	5	2.3	35	12	Q80477	human	immun
541	35	10	Q39519	nicotiana t	614	5	5	2.3	35	12	Q80591	human	immun
542	35	12	Q76251	human	immun	615	5	2.3	35	12	Q70780	human	immun
543	35	12	Q76374	human	immun	616	5	2.3	35	12	Q70781	human	immun
544	35	12	Q76374	human	immun	617	5	2.3	35	12	Q70782	human	immun
545	35	12	Q7629	human	immun	618	5	2.3	35	12	Q70783	human	immun
546	35	12	Q7644	human	immun	619	5	2.3	35	12	Q70810	human	immun
547	35	12	O40580	human	immun	620	5	2.3	35	12	Q70811	human	immun
548	35	12	O40586	human	immun	621	5	2.3	35	12	Q78492	human	immun
549	35	12	Q70414	human	immun	622	5	2.3	35	12	Q78493	human	immun
550	35	12	Q70427	human	immun	623	5	2.3	35	12	Q78493	human	immun
551	35	12	Q76296	human	immun	624	5	2.3	35	12	O10869	human	immun
552	35	12	Q76369	human	immun	625	5	2.3	35	12	O90503	human	immun
553	35	12	Q78035	human	immun	626	5	2.3	35	12	O90513	human	immun
554	35	12	Q78039	human	immun	627	5	2.3	35	12	O90525	human	immun
555	35	12	Q78043	human	immun	628	5	2.3	35	12	O91349	human	immun
556	35	12	Q78055	human	immun	629	5	2.3	35	12	O91350	human	immun
557	35	12	Q78059	human	immun	630	5	2.3	35	12	O91351	human	immun
558	35	12	Q78094	human	immun	631	5	2.3	35	12	O91352	human	immun
559	35	12	Q78102	human	immun	632	5	2.3	35	12	O91353	human	immun
560	35	12	Q78107	human	immun	633	5	2.3	35	12	O91430	human	immun
561	35	12	Q78110	human	immun	634	5	2.3	35	12	O91433	human	immun
562	35	12	Q78198	human	immun	635	5	2.3	35	12	O91434	human	immun
563	35	12	Q78201	human	immun	636	5	2.3	35	12	O91435	human	immun
564	35	12	Q80012	human	immun	637	5	2.3	35	12	O91436	human	immun
565	35	12	Q80465	human	immun	638	5	2.3	35	12	O91437	human	immun
566	35	12	Q80476	human	immun	639	5	2.3	35	12	O91439	human	immun
567	35	12	Q80478	human	immun	640	5	2.3	35	12	O91440	human	immun
568	35	12	Q80519	human	immun	641	5	2.3	35	12	O91441	human	immun
569	35	12	Q80590	human	immun	642	5	2.3	35	12	O91442	human	immun
570	35	12	Q80592	human	immun	643	5	2.3	35	12	O91443	human	immun
571	35	12	Q80603	human	immun	644	5	2.3	35	12	O91444	human	immun
572	35	12	Q80603	human	immun	645	5	2.3	35	12	O91445	human	immun
573	35	12	Q71189	human	immun	646	5	2.3	35	12	O91446	human	immun
574	35	12	Q71190	human	immun	647	5	2.3	35	12	O91447	human	immun
575	35	12	Q71191	human	immun	648	5	2.3	35	12	Q76276	human	immun
576	35	12	Q77996	human	immun	649	5	2.3	35	12	Q76276	human	immun
577	35	12	Q78026	human	immun	650	5	2.3	35	12	Q7931	human	immun
578	35	12	Q78046	human	immun	651	5	2.3	35	12	Q79420	human	immun
579	35	12	Q78048	human	immun	652	5	2.3	35	12	Q76295	human	immun
580	35	12	Q78051	human	immun	653	5	2.3	35	12	Q76295	human	immun
581	35	12	Q78056	human	immun	654	5	2.3	35	12	Q70232	human	immun
582	35	12	Q78065	human	immun	655	5	2.3	35	12	Q70232	human	immun
583	35	12	Q78067	human	immun	656	5	2.3	35	12	Q70232	human	immun
584	35	12	Q78071	human	immun	657	5	2.3	35	12	Q9WCX4	human	immun
585	35	12	Q78072	human	immun	658	5	2.3	35	12	Q9WCX4	human	immun
586	35	12	Q78075	human	immun	659	5	2.3	35	12	Q9WCW6	human	immun
587	35	12	Q78079	human	immun	660	5	2.3	35	12	Q9WCW6	human	immun
588	35	12	Q78081	human	immun	661	5	2.3	35	12	Q9WCW5	human	immun
589	35	12	Q78082	human	immun	662	5	2.3	35	12	Q9W8V8	human	immun
590	35	12	Q78083	human	immun	663	5	2.3	35	12	Q9W8V8	human	immun
591	35	12	Q78084	human	immun	664	5	2.3	35	12	O10487	human	immun
592	35	12	Q78087	human	immun	665	5	2.3	35	12	O10487	human	immun
593	35	12	Q78091	human	immun	666	5	2.3	35	12	O24580	zea mays	immun
594	35	12	Q78093	human	immun	667	5	2.3	35	12	O24580	zea mays	immun
595	35	12	Q78096	human	immun	668	5	2.3	35	12	Q72317	human	immun
596	35	12	Q78097	human	immun	669	5	2.3	35	12	Q72317	human	immun
597	35	12	Q78103	human	immun	670	5	2.3	35	12	Q72845	human	immun
598	35	12	Q78104	human	immun	671	5	2.3	35	12	Q72846	human	immun
599	35	12	Q78105	human	immun	672	5	2.3	35	12	Q72847	human	immun
600	35	12	Q78106	human	immun	673	5	2.3	35	12	Q72847	human	immun
601	35	12	Q78111	human	immun	674	5	2.3	35	12	Q72854	human	immun
602	35	12	Q78122	human	immun	675	5	2.3	35	12	Q72854	human	immun
603	35	12	Q78136	human	immun	676	5	2.3	35	12	Q72859	human	immun
			Q78139	human	immun						Q72869	human	immun
											Q72870	human	immun
											Q72870	human	immun
											P97550	rattus norv	immun
											Q84307	human	immun
											Q84308	human	immun
											Q92307	human	immun
											Q92307	human	immun
											Q92315	human	immun
											Q92318	human	immun
											Q9WJb0	human	immun
											Q9wa44	human	immun
											Q62307	mus musculus	immun
											Q71191	grapevine	immun
											Q9WCW2	human	immun
											O50275	rhodococcus	immun

677	5	2.3	46	10	Q40923	O40923	pseudotsuga	750	5	2.3	54	12	Q69801	Q69801	human	immun
678	5	2.3	46	12	Q72529	Q72529	human	751	5	2.3	54	12	Q69802	Q69802	human	immun
679	5	2.3	46	12	Q82401	Q82401	human t-cel	752	5	2.3	54	12	Q69804	Q69804	human	immun
680	5	2.3	46	12	Q82402	Q82402	human t-cel	753	5	2.3	54	12	Q69805	Q69805	human	immun
681	5	2.3	46	12	Q82403	Q82403	human t-cel	754	5	2.3	54	12	Q69806	Q69806	human	immun
682	5	2.3	46	12	Q82404	Q82404	human t-cel	755	5	2.3	54	12	Q69811	Q69811	human	immun
683	5	2.3	46	12	Q82407	Q82407	human t-cel	756	5	2.3	54	12	Q69821	Q69821	human	immun
684	5	2.3	46	12	Q82407	Q82407	human t-cel	757	5	2.3	54	12	Q69824	Q69824	human	immun
685	5	2.3	48	11	Q63885	Q63885	mus musculus	758	5	2.3	54	12	Q69827	Q69827	human	immun
686	5	2.3	48	12	Q91284	Q91284	human	759	5	2.3	54	12	Q69830	Q69830	human	immun
687	5	2.3	49	2	Q50263	Q50263	phytoplasma	760	5	2.3	54	12	Q69831	Q69831	human	immun
688	5	2.3	49	12	Q9YL02	Q9YL02	human	761	5	2.3	54	12	Q70284	Q70284	human	immun
689	5	2.3	50	3	P89114	P89114	saccharomyc	762	5	2.3	54	12	Q76486	Q76486	human	immun
690	5	2.3	50	12	Q9WMN9	Q9WMN9	human polio	763	5	2.3	54	12	Q76491	Q76491	human	immun
691	5	2.3	50	12	Q9WMN8	Q9WMN8	human polio	764	5	2.3	54	12	Q76492	Q76492	human	immun
692	5	2.3	51	11	Q88660	Q88660	rattus norv	765	5	2.3	54	12	Q76494	Q76494	human	immun
693	5	2.3	51	12	Q37475	Q37475	human	766	5	2.3	54	12	Q76498	Q76498	human	immun
694	5	2.3	52	5	Q97152	Q97152	drosophila	767	5	2.3	54	12	Q76499	Q76499	human	immun
695	5	2.3	53	2	Q57170	Q57170	clostridium	768	5	2.3	54	12	Q76500	Q76500	human	immun
696	5	2.3	53	4	Q9Y6W9	Q9Y6W9	homo sapien	769	5	2.3	54	12	Q76501	Q76501	human	immun
697	5	2.3	53	10	Q41894	Q41894	zea mays (m	770	5	2.3	54	12	Q76502	Q76502	human	immun
698	5	2.3	53	12	Q69812	Q69812	human	771	5	2.3	54	12	Q76503	Q76503	human	immun
699	5	2.3	53	12	Q69828	Q69828	human	772	5	2.3	54	12	Q76504	Q76504	human	immun
700	5	2.3	53	12	Q76523	Q76523	human	773	5	2.3	54	12	Q76516	Q76516	human	immun
701	5	2.3	53	12	Q76546	Q76546	human	774	5	2.3	54	12	Q76527	Q76527	human	immun
702	5	2.3	53	12	Q97824	Q97824	human	775	5	2.3	54	12	Q69798	Q69798	human	immun
703	5	2.3	53	12	Q97826	Q97826	human	776	5	2.3	54	12	Q69803	Q69803	human	immun
704	5	2.3	53	12	Q97831	Q97831	human	777	5	2.3	54	12	Q69808	Q69808	human	immun
705	5	2.3	53	12	Q97835	Q97835	human	778	5	2.3	54	12	Q69810	Q69810	human	immun
706	5	2.3	53	12	Q97843	Q97843	human	779	5	2.3	54	12	Q70278	Q70278	human	immun
707	5	2.3	53	12	Q97844	Q97844	human	780	5	2.3	54	12	Q70279	Q70279	human	immun
708	5	2.3	53	12	Q97844	Q97844	human	781	5	2.3	54	12	Q70280	Q70280	human	immun
709	5	2.3	53	12	Q97949	Q97949	human	782	5	2.3	54	12	Q70281	Q70281	human	immun
710	5	2.3	53	12	Q97950	Q97950	human	783	5	2.3	54	12	Q70282	Q70282	human	immun
711	5	2.3	53	12	Q97954	Q97954	human	784	5	2.3	54	12	Q76488	Q76488	human	immun
712	5	2.3	53	12	Q97957	Q97957	human	785	5	2.3	54	12	Q76489	Q76489	human	immun
713	5	2.3	53	12	Q97958	Q97958	human	786	5	2.3	54	12	Q76490	Q76490	human	immun
714	5	2.3	53	12	Q97962	Q97962	human	787	5	2.3	54	12	Q76495	Q76495	human	immun
715	5	2.3	53	12	Q97967	Q97967	human	788	5	2.3	54	12	Q76505	Q76505	human	immun
716	5	2.3	53	12	Q97968	Q97968	human	789	5	2.3	54	12	Q76506	Q76506	human	immun
717	5	2.3	53	12	Q97973	Q97973	human	790	5	2.3	54	12	Q76517	Q76517	human	immun
718	5	2.3	53	12	Q97820	Q97820	human	791	5	2.3	54	12	Q76526	Q76526	human	immun
719	5	2.3	53	12	Q97821	Q97821	human	792	5	2.3	54	12	Q37466	Q37466	human	immun
720	5	2.3	53	12	Q97822	Q97822	human	793	5	2.3	54	12	Q37467	Q37467	human	immun
721	5	2.3	53	12	Q97823	Q97823	human	794	5	2.3	56	10	Q41836	Q41836	zea mays (m	
722	5	2.3	53	12	Q97825	Q97825	human	795	5	2.3	56	12	Q37455	Q37455	human	immun
723	5	2.3	53	12	Q97828	Q97828	human	796	5	2.3	56	12	Q37458	Q37458	human	immun
724	5	2.3	53	12	Q97829	Q97829	human	797	5	2.3	56	12	Q37462	Q37462	human	immun
725	5	2.3	53	12	Q97833	Q97833	human	798	5	2.3	56	12	Q37465	Q37465	human	immun
726	5	2.3	53	12	Q97836	Q97836	human	799	5	2.3	56	12	Q37469	Q37469	human	immun
727	5	2.3	53	12	Q97837	Q97837	human	800	5	2.3	56	12	Q37471	Q37471	human	immun
728	5	2.3	53	12	Q97840	Q97840	human	801	5	2.3	56	12	Q37472	Q37472	human	immun
729	5	2.3	53	12	Q97841	Q97841	human	802	5	2.3	56	12	Q37476	Q37476	human	immun
730	5	2.3	53	12	Q97952	Q97952	human	803	5	2.3	56	12	Q37478	Q37478	human	immun
731	5	2.3	53	12	Q97953	Q97953	human	804	5	2.3	56	12	Q37562	Q37562	human	immun
732	5	2.3	53	12	Q97955	Q97955	human	805	5	2.3	57	12	Q37791	Q37791	human	immun
733	5	2.3	53	12	Q97956	Q97956	human	806	5	2.3	57	12	Q37804	Q37804	human	immun
734	5	2.3	53	12	Q97960	Q97960	human	807	5	2.3	57	12	Q37806	Q37806	human	immun
735	5	2.3	53	12	Q97961	Q97961	human	808	5	2.3	57	12	Q37810	Q37810	human	immun
736	5	2.3	53	12	Q97963	Q97963	human	809	5	2.3	57	12	Q37811	Q37811	human	immun
737	5	2.3	53	12	Q97964	Q97964	human	810	5	2.3	57	12	Q37816	Q37816	human	immun
738	5	2.3	53	12	Q97969	Q97969	human	811	5	2.3	57	12	Q37817	Q37817	human	immun
739	5	2.3	53	12	Q97970	Q97970	human	812	5	2.3	58	12	Q85772	Q85772	human	immun
740	5	2.3	53	12	Q97971	Q97971	human	813	5	2.3	58	12	Q85782	Q85782	human	immun
741	5	2.3	53	12	Q97972	Q97972	human	814	5	2.3	59	12	P90112	P90112	human	immun
742	5	2.3	53	12	Q37554	Q37554	human	815	5	2.3	59	12	O40613	O40613	hepatitis c	
743	5	2.3	53	12	Q37556	Q37556	human	816	5	2.3	59	12	O40614	O40614	hepatitis c	
744	5	2.3	53	12	Q37557	Q37557	human	817	5	2.3	59	12	O37563	O37563	human	immun
745	5	2.3	54	2	Q69872	Q69872	streptomyce	818	5	2.3	60	10	O49462	O49462	arabidopsis	
746	5	2.3	54	12	Q76143	Q76143	human	819	5	2.3	60	12	Q85738	Q85738	human	immun
747	5	2.3	54	12	Q76300	Q76300	human	820	5	2.3	60	12	Q85742	Q85742	human	immun
748	5	2.3	54	12	Q69799	Q69799	human	821	5	2.3	60	12	Q85752	Q85752	human	immun
749	5	2.3	54	12	Q69800	Q69800	human	822	5	2.3	60	12	Q85776	Q85776	human	immun

823	60	12	Q85753	Q85753 human immun	896	5	2.3	63	12	Q37508	Q37508 human immun
824	60	12	Q85754	Q85754 human immun	897	5	2.3	63	12	Q37509	Q37509 human immun
825	60	12	Q37794	Q37794 human immun	898	5	2.3	63	12	Q37510	Q37510 human immun
826	60	12	Q37795	Q37795 human immun	899	5	2.3	63	12	Q37511	Q37511 human immun
827	60	12	Q37796	Q37796 human immun	900	5	2.3	63	12	Q37512	Q37512 human immun
828	61	1	Q9VE88	Q9VE88 aeropyrum p	901	5	2.3	63	12	Q37513	Q37513 human immun
829	61	12	Q85739	Q85739 human immun	902	5	2.3	63	12	Q37514	Q37514 human immun
830	61	12	Q85773	Q85773 human immun	903	5	2.3	63	12	Q37515	Q37515 human immun
831	61	12	Q85777	Q85777 human immun	904	5	2.3	63	12	Q37516	Q37516 human immun
832	61	12	Q85779	Q85779 human immun	905	5	2.3	63	12	Q37517	Q37517 human immun
833	61	12	Q85812	Q85812 human immun	906	5	2.3	63	12	Q37518	Q37518 human immun
834	61	12	Q85813	Q85813 human immun	907	5	2.3	63	12	Q37519	Q37519 human immun
835	61	12	Q85817	Q85817 human immun	908	5	2.3	63	12	Q37520	Q37520 human immun
836	61	12	Q85823	Q85823 human immun	909	5	2.3	63	12	Q37521	Q37521 human immun
837	61	12	Q85773	Q85773 human immun	910	5	2.3	63	12	Q37522	Q37522 human immun
838	61	12	Q85810	Q85810 human immun	911	5	2.3	63	12	Q37523	Q37523 human immun
839	61	12	Q10938	Q10938 human immun	912	5	2.3	63	12	Q37524	Q37524 human immun
840	61	12	Q37807	Q37807 human immun	913	5	2.3	63	12	Q37525	Q37525 human immun
841	61	12	Q37809	Q37809 human immun	914	5	2.3	63	12	Q37527	Q37527 human immun
842	62	4	Q14311	Q14311 homo sapien	915	5	2.3	63	12	Q37530	Q37530 human immun
843	62	4	Q16041	Q16041 homo sapien	916	5	2.3	63	12	Q37533	Q37533 human immun
844	62	6	Q9XSB3	Q9XSB3 canis famil	917	5	2.3	63	12	Q37534	Q37534 human immun
845	62	9	Q9XJM2	Q9XJM2 bacterioph	918	5	2.3	63	12	Q37535	Q37535 human immun
846	62	12	Q85745	Q85745 human immun	919	5	2.3	63	12	Q37536	Q37536 human immun
847	62	12	Q85748	Q85748 human immun	920	5	2.3	63	12	Q37537	Q37537 human immun
848	62	12	Q85784	Q85784 human immun	921	5	2.3	63	12	Q37538	Q37538 human immun
849	62	12	Q85785	Q85785 human immun	922	5	2.3	63	12	Q37539	Q37539 human immun
850	62	12	Q37457	Q37457 human immun	923	5	2.3	63	12	Q37540	Q37540 human immun
851	62	12	Q37559	Q37559 human immun	924	5	2.3	63	12	Q37541	Q37541 human immun
852	62	12	Q9Y296	Q9Y296 human immun	925	5	2.3	63	12	Q37542	Q37542 human immun
853	62	12	Q9Y292	Q9Y292 human immun	926	5	2.3	63	12	Q37543	Q37543 human immun
854	62	12	Q9XZ84	Q9XZ84 human immun	927	5	2.3	63	12	Q37544	Q37544 human immun
855	63	2	Q31968	Q31968 bacillus su	928	5	2.3	63	12	Q37545	Q37545 human immun
856	63	7	Q31236	Q31236 mus muscucu	929	5	2.3	63	12	Q37546	Q37546 human immun
857	63	9	Q64052	Q64052 bacterioph	930	5	2.3	63	12	Q37547	Q37547 human immun
858	63	12	Q75186	Q75186 human immun	931	5	2.3	63	12	Q37548	Q37548 human immun
859	63	12	Q85747	Q85747 human immun	932	5	2.3	63	12	Q37550	Q37550 human immun
860	63	12	Q85786	Q85786 human immun	933	5	2.3	63	12	Q37551	Q37551 human immun
861	63	12	Q85809	Q85809 human immun	934	5	2.3	63	12	Q37552	Q37552 human immun
862	63	12	Q85818	Q85818 human immun	935	5	2.3	63	12	Q37553	Q37553 human immun
863	63	12	Q85822	Q85822 human immun	936	5	2.3	63	12	Q37555	Q37555 human immun
864	63	12	Q37454	Q37454 human immun	937	5	2.3	63	12	Q37558	Q37558 human immun
865	63	12	Q37459	Q37459 human immun	938	5	2.3	63	12	Q37560	Q37560 human immun
866	63	12	Q37460	Q37460 human immun	939	5	2.3	63	12	Q37565	Q37565 human immun
867	63	12	Q37461	Q37461 human immun	940	5	2.3	63	12	Q37571	Q37571 human immun
868	63	12	Q37463	Q37463 human immun	941	5	2.3	63	12	Q37572	Q37572 human immun
869	63	12	Q37464	Q37464 human immun	942	5	2.3	63	12	Q37574	Q37574 human immun
870	63	12	Q37468	Q37468 human immun	943	5	2.3	63	12	Q37575	Q37575 human immun
871	63	12	Q37474	Q37474 human immun	944	5	2.3	63	12	Q37576	Q37576 human immun
872	63	12	Q37479	Q37479 human immun	945	5	2.3	63	12	Q37578	Q37578 human immun
873	63	12	Q37482	Q37482 human immun	946	5	2.3	63	12	Q37579	Q37579 human immun
874	63	12	Q37483	Q37483 human immun	947	5	2.3	63	12	Q37580	Q37580 human immun
875	63	12	Q37484	Q37484 human immun	948	5	2.3	63	12	Q37582	Q37582 human immun
876	63	12	Q37485	Q37485 human immun	949	5	2.3	63	12	Q37583	Q37583 human immun
877	63	12	Q37486	Q37486 human immun	950	5	2.3	63	12	Q37587	Q37587 human immun
878	63	12	Q37488	Q37488 human immun	951	5	2.3	63	12	Q37589	Q37589 human immun
879	63	12	Q37489	Q37489 human immun	952	5	2.3	63	12	Q37590	Q37590 human immun
880	63	12	Q37490	Q37490 human immun	953	5	2.3	63	12	Q37591	Q37591 human immun
881	63	12	Q37492	Q37492 human immun	954	5	2.3	63	12	Q37592	Q37592 human immun
882	63	12	Q37493	Q37493 human immun	955	5	2.3	63	12	Q37593	Q37593 human immun
883	63	12	Q37494	Q37494 human immun	956	5	2.3	63	12	Q37594	Q37594 human immun
884	63	12	Q37495	Q37495 human immun	957	5	2.3	63	12	Q37595	Q37595 human immun
885	63	12	Q37496	Q37496 human immun	958	5	2.3	63	12	Q37596	Q37596 human immun
886	63	12	Q37498	Q37498 human immun	959	5	2.3	63	12	Q37599	Q37599 human immun
887	63	12	Q37499	Q37499 human immun	960	5	2.3	63	12	Q37617	Q37617 human immun
888	63	12	Q37500	Q37500 human immun	961	5	2.3	63	12	Q37771	Q37771 human immun
889	63	12	Q37501	Q37501 human immun	962	5	2.3	63	12	Q37772	Q37772 human immun
890	63	12	Q37502	Q37502 human immun	963	5	2.3	63	12	Q37773	Q37773 human immun
891	63	12	Q37503	Q37503 human immun	964	5	2.3	63	12	Q37774	Q37774 human immun
892	63	12	Q37504	Q37504 human immun	965	5	2.3	63	12	Q37775	Q37775 human immun
893	63	12	Q37505	Q37505 human immun	966	5	2.3	63	12	Q37776	Q37776 human immun
894	63	12	Q37506	Q37506 human immun	967	5	2.3	63	12	Q37777	Q37777 human immun
895	63	12	Q37507	Q37507 human immun	968	5	2.3	63	12	Q37778	Q37778 human immun

969 5 2.3 63 12 037785 037785 human immun
 970 5 2.3 63 12 037788 037788 human immun
 971 5 2.3 63 12 037793 037793 human immun
 972 5 2.3 63 12 037823 037823 human immun
 973 5 2.3 63 12 037841 037841 human immun
 974 5 2.3 63 12 037842 037842 human immun
 975 5 2.3 63 12 037843 037843 human immun
 976 5 2.3 63 12 037844 037844 human immun
 977 5 2.3 63 12 037845 037845 human immun
 978 5 2.3 63 12 037846 037846 human immun
 979 5 2.3 63 12 037847 037847 human immun
 980 5 2.3 63 12 037848 037848 human immun
 981 5 2.3 63 12 037850 037850 human immun
 982 5 2.3 63 12 037851 037851 human immun
 983 5 2.3 63 12 037852 037852 human immun
 984 5 2.3 63 12 037858 037858 human immun
 985 5 2.3 63 12 037859 037859 human immun
 986 5 2.3 63 12 037860 037860 human immun
 987 5 2.3 63 12 037864 037864 human immun
 988 5 2.3 63 12 037865 037865 human immun
 989 5 2.3 63 12 037869 037869 human immun
 990 5 2.3 63 12 037872 037872 human immun
 991 5 2.3 63 12 037873 037873 human immun
 992 5 2.3 63 12 037874 037874 human immun
 993 5 2.3 63 12 037875 037875 human immun
 994 5 2.3 63 12 037876 037876 human immun
 995 5 2.3 63 12 037877 037877 human immun
 996 5 2.3 63 12 085787 085787 human immun
 997 5 2.3 64 2 092667 092667 chlamydia t
 998 5 2.3 64 8 032956 032956 pinus thumb
 999 5 2.3 64 8 035890 035890 simuliun ny
 1000 5 2.3 64 11 063005 063005 rattus norv

ALIGNMENTS

RESULT 1
 ID P90743 PRELIMINARY; PRT; 341 AA.
 AC P90743;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
 DE C10C6.3 PROTEIN.
 GN C10C6.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA WHITE S.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTINORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL Nature 368:32-38(1994).
 DR EMBL; 283217; CAB05680.1;
 SQ SEQUENCE 341 AA; 37185 MW; 30ED368E CRC32;

Query Match 3.6%; Score 8; DB 5; Length 341;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 113 VQVEGEPD 120
 DB 45 VQVEGEPD 52
 |||||
 |||||
 RESULT 2
 OS Treponema pallidum.
 ID 083284 PRELIMINARY; PRT; 448 AA.
 AC 083284;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 51.2 KD PROTEIN.
 GN TP0260.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98332770.
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
 RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
 RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
 RA VENTER J.C.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A5001207; AAC65253.1;
 DR TIGR; TP0260;
 KW Hypothetical protein.
 SQ SEQUENCE 448 AA; 51247 MW; 62486D07 CRC32;

Query Match 3.6%; Score 8; DB 2; Length 448;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 QFTDEAGR 62
 DB 287 QFTDEAGR 294
 |||||
 |||||
 RESULT 3
 ID Q20943 PRELIMINARY; PRT; 1494 AA.
 AC Q20943;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE SIMILAR TO THE HUMAN MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN.
 GN F57C12.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.

RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BORTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON A., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER M., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RUFEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIRRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA FAVELLO T.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA WATERSTON R.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U41554; AA83299.1; -;
 DR HSSP; P13569; INBD.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 DR PFAM; PF00654; ABC_membrane; 2.
 DR PFAM; PF00005; ABC_tran; 2.
 KW ATP-binding; Transport.
 SQ SEQUENCE 1494 AA; 166742 MW; 628DF55C CRC32;

Query Match 3.6%; Score 8; DB 5; Length 1494;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 44 ILFNKPYD 51
 |||||
 Db 651 ILFNKPYD 658

RESULT 4
 ID Q94137 PRELIMINARY; PRT; 1525 AA.
 AC Q94137;
 DT 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
 DE MULTIDRUG RESISTANCE RELATED PROTEIN 2.
 GN MRP-2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BROEKS A., GERRARD B., ALLIKMETS A., DEAN M., PLASTERK R.H.A.;
 RL EMBO J. 0:0-0(0).
 DR EMBL; U66261; AA807022.1; -;
 DR HSSP; P13569; INBD.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 DR PFAM; PF00654; ABC_membrane; 2.
 DR PFAM; PF00005; ABC_tran; 2.
 KW ATP-binding; Transport.
 SQ SEQUENCE 1525 AA; 170216 MW; DC104A75 CRC32;

Query Match 3.6%; Score 8; DB 5; Length 1525;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 44 ILFNKPYD 51
 |||||
 Db 696 ILFNKPYD 703

RESULT 5
 Q38044
 ID Q38044 PRELIMINARY; PRT; 66 AA.
 AC Q38044;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
 DE EXCISIONASE.
 OS Bacteriophage phi-11.
 OC Viruses.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90236919.
 RA YE Z.H., BURANEN S.L., LEE C.Y.;
 RT "Sequence analysis and comparison of int and xis genes from
 RT staphylococcal bacteriophages L54a and phi 11.";
 RL J. Bacteriol. 172:2568-2575(1990).
 DR EMBL; M34832; AAA32197.1; -;
 SQ SEQUENCE 66 AA; 7639 MW; 3BB115E4 CRC32;

Query Match 3.2%; Score 7; DB 9; Length 66;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 204 YTLDLSLA 210
 |||||
 Db 54 YTLDLSLA 60

RESULT 6
 ID Q51543 PRELIMINARY; PRT; 288 AA.
 AC Q51543;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
 DE POBR AND POBA GENES.
 GN POBR.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 OC Pseudomonas.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PAOIC;
 RA ENTSCHE B., SQUIRE L., WICKS R.E.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X76994; CAA54301.1; -;
 DR PFAM; PF00165; HTH_2; 1.
 DR PRINTS; PR00032; HTHARAC.
 SQ SEQUENCE 288 AA; 32176 MW; DDED4F70 CRC32;

Query Match 3.2%; Score 7; DB 2; Length 288;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 113 VQVEGEP 119
 |||||
 Db 63 VQVEGEP 69

RESULT 7
 ID P72806 PRELIMINARY; PRT; 312 AA.
 AC P72806;
 DT 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
 DE HYPOTHETICAL 35.6 KD PROTEIN.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARDO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90900; BAA16821.1; -.
DR PFAM; PF00535; Glycos_transf_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 312 AA; 35576 MW; EB3AE5A0 CRC32;

Query Match 3.2%; Score 7; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 VOGVTAA 77
DB 298 VOGVTAA 304

RESULT 8
QY05938 PRELIMINARY; PRT; 352 AA.
ID C05938;
AC C05938;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE LIPASE HELPER PROTEIN.
GN LPWB.
OS Pseudomonas wisconsinensis.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LMG P-15151;
RA HAZBON M.H., DUFEL H., CORNELIS P., JAEGER K.E.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN-LMG P-15151;
RA ANDRE C., CHARMOILLE L., CORNELIS P., HAZBON M.H.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88907; AAB53648.1; -.
SQ SEQUENCE 352 AA; 39574 MW; 149A3B5C CRC32;

Query Match 3.2%; Score 7; DB 2; Length 352;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 YTLDSLA 210
DB 203 YTLDSLA 209

RESULT 9
QY179 PRELIMINARY; PRT; 356 AA.
ID QY179;
AC QY179;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE SUBTILISIN-LIKE PROTEASE (EC 3.4.21.14).

```

```

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE AMINOPHOSPHOLIPID TRANSLOCASE (FRAGMENT).
GN APLT-1.
OS Trichomonas vaginalis.
OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC30325;
RA SHAH P.H., LUSHBAUGH W.B., FINLEY R.W., LI C.L., MOATE M.E.,
RA MEADE J.C.;
RT "Characterization of a family of p-type ATPases in Trichomonas
RT vaginalis.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF145277; AAD37686.1; -.
FT NON_TER 1
FT NON_TER 356
SQ SEQUENCE 356 AA; 40036 MW; 8B6B88D9 CRC32;

Query Match 3.2%; Score 7; DB 5; Length 356;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 EGLLVLT 91
DB 127 EGLLVLT 133

RESULT 10
QY2MQ3 PRELIMINARY; PRT; 406 AA.
ID Q92MQ3;
AC Q92MQ3;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE PUTATIVE.
GN JHP0165.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J99;
RX MEDLINE; 99120557.
RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA TUMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA TRUST T.J.;
RT "Genomic-sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001455; AAD05746.1; -.
SQ SEQUENCE 406 AA; 47908 MW; EC11F71A CRC32;

Query Match 3.2%; Score 7; DB 2; Length 406;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 ANGEWRD 216
DB 356 ANGEWRD 362

RESULT 11
QY0206 PRELIMINARY; PRT; 416 AA.
ID Q00206;
AC Q00206;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SUBTILISIN-LIKE PROTEASE (EC 3.4.21.14).

```

GN pepD.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
 OC Eurotiales; Trichocomaceae; anamorphic Trichocomaceae; Aspergillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-N400;
 RX MEDLINE; 94156181.
 RA JARAI G., BUXTON F.P.;
 RT "Cloning and characterization of the pepD gene of Aspergillus niger
 which codes for a subtilisin-like protease.";
 RL Gene 139:51-57(1994).
 DR EMBL; L19059; AAA32703.1; -.
 DR HSP; P06873; IBJR.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR PRAM; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 KW Protease; Hydrolase.
 SQ SEQUENCE 416 AA; 43899 MW; 43279E08 CRC32;

Query Match 3.2%; Score 7; DB 3; Length 416;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LPAGIER 145
 |||||
 DB 75 LPAGIER 81

RESULT 12
 ID O86823 PRELIMINARY; PRT; 444 AA.
 AC O86823;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE PUTATIVE 4-AMINOBTYRATE AMINOTRANSFERASE.
 GN GABT.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MURPHY L., HARRIS D.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE; 97000351,
 RA REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
 RA KINASHI H., HOPWOOD D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL031225; CAA20213.1; -.
 DR HSP; P16932; 1DGE.
 DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
 DR PRAM; PF00202; aminotran_3; 1.
 KW Transferase; Aminotransferase.
 SQ SEQUENCE 444 AA; 46479 MW; 24050EB1 CRC32;

Query Match 3.2%; Score 7; DB 2; Length 444;
 Best Local Similarity 100.0%; Pred. No. 54;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 85 EGLVLT 91
 |||||
 DB 403 EGLVLT 409
 RESULT 13
 ID O25572 PRELIMINARY; PRT; 514 AA.
 AC O25572;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 58.9 KD PROTEIN.
 GN HP0914.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26695;
 RX MEDLINE; 97394467.
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
 RA FREISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori.";
 RL Nature 388:539-547(1997).
 DR EMBL; AS000601; AAD07966.1; -.
 DR TIGR; HP0914; -.
 KW Hypothetical protein.
 SQ SEQUENCE 514 AA; 58854 MW; 7317BE1E CRC32;

Query Match 3.2%; Score 7; DB 2; Length 514;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 RNVTLN 134
 |||||
 DB 326 RNVTLN 332
 RESULT 14
 ID O92KT5 PRELIMINARY; PRT; 514 AA.
 AC O92KT5;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE PUTATIVE OUTER MEMBRANE PROTEIN.
 GN JHP0850.
 OS Helicobacter pylori J99.
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-J99;
 RX MEDLINE; 99120557.
 RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
 RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
 RA TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
 RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
 RA TRUST T.J.;
 RT "Genomic-sequence comparison of two unrelated isolates of the human
 gastric pathogen Helicobacter pylori.";

RL Nature 397:176-180(1999).
DR EMBL; AE001514; AAD08428.1; -.
SQ SEQUENCE 514 AA; 58892 MW; 198C926D CRC32;

Query Match 3.2%; Score 7; DB 2; Length 514;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRGVTLN 134
|||||
DB 326 RRGVTLN 332

RESULT 15
O14407
ID O14407 PRELIMINARY; PRT; 710 AA.
AC O14407;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE ROPY-2.
GN ROPY-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC Sordariales; Sordariaceae; Neurospora.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-74-OR23-1VA;
RA VIERULA J.P., MAIS J.M.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U23425; AAB68457.1; -.
SQ SEQUENCE 710 AA; 77817 MW; DBF775EA CRC32;

Query Match 3.2%; Score 7; DB 3; Length 710;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 LAKLRNG 130
|||||
DB 347 LAKLRNG 353

Search completed: June 10, 2000, 11:30:40
Job time: 335 sec


```
Ratio: 1.000          Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-123-936-182
Align seg 1/1 to: US-08-123-936-182 from: 1 to: 46

98 AlaArgLeuThrGlnProGly 104
|||||
3 GCCCGGCTGACTCAGCCGGG 23

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-475-228A-182

seq_documentation_block:
; Sequence 182, Application US/08475228A
; Patent No. 5869241
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,228A
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human hsp70B gene
US-08-475-228A-182
```

```
alignment_scores:
Quality: 7.00          Length: 7
Ratio: 1.000          Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-475-228A-182
Align seg 1/1 to: US-08-475-228A-182 from: 1 to: 46

98 AlaArgLeuThrGlnProGly 104
|||||
3 GCCCGGCTGACTCAGCCGGG 23

seq_name: /cgn2_6/ptodata/1/lna/6_COMB.seq:US-08-482-080A-182

seq_documentation_block:
; Sequence 182, Application US/08482080A
; Patent No. 6010849
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,080A
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/171,389
; FILING DATE: 20-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brady, John F.
; REGISTRATION NUMBER: 39,118
; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 324-0880
; TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
```

```
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human hsp70B gene
US-08-482-080A-182

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-482-080A-182
..
Align seg 1/1 to: US-08-482-080A-182 from: 1 to: 46

98 AlaArgLeuThrGlnProGly 104
|||||
3 GCCCGCTGACTCAGCCGGG 23

seq_name: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:PCT-US93-12388-182

seq_documentation_block:
; Sequence 182, Application PC/TUS9312388
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12388
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human hsp70B gene
PCT-US93-12388-182

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
```

```
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-252-691-7056 x PCT-US93-12388-182
..
Align seg 1/1 to: PCT-US93-12388-182 from: 1 to: 46

98 AlaArgLeuThrGlnProGly 104
|||||
3 GCCCGCTGACTCAGCCGGG 23

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-468-819-73

seq_documentation_block:
; Sequence 73, Application US/08468819
; Patent No. 5871723
; GENERAL INFORMATION:
; APPLICANT: Strieter, Robert M.
; APPLICANT: Polverini, Peter J.
; APPLICANT: Kunkel, Steven L.
; TITLE OF INVENTION: CXC Chemokines as Regulators of
; TITLE OF INVENTION: Angiogenesis
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,819
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UMIC:003/HYL
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7477
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-468-819-73

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-468-819-73/rev
..
Align seg 1/1 to reverse of: US-08-468-819-73 from: 1 to: 654

157 ProProfileArgGluArgLys 163
|||||
196 CCACCTATCAGGGAACGGAAG 176

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-468-819-75
```

```
seq_documentation_block:
; Sequence 75, Application US/08468819
; Patent No. 5871723
; GENERAL INFORMATION:
; APPLICANT: Strieter, Robert M.
; APPLICANT: Polverini, Peter J.
; APPLICANT: Kunkel, Steven L.
; TITLE OF INVENTION: CXC Chemokines as Regulators of
; TITLE OF INVENTION: Angiogenesis
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,819
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UMC:003/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7477
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; US-08-468-819-75

alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
; US-09-252-691-7056 x US-08-468-819-75/rev ..
; Align seg 1/1 to reverse of: US-08-468-819-75 from: 1 to: 654

157 ProProileArgGluArgLys 163
|||||
196 CCACCTATCAGGAACGGAAG 176

seq_name: /cgn2.6/ptodata/1/lna/5D_COMB.seq:US-08-997-080-40

seq_documentation_block:
; Sequence 40, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
```

```
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-997-080-40

alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
; US-09-252-691-7056 x US-08-997-080-40/rev ..
; Align seg 1/1 to reverse of: US-08-997-080-40 from: 1 to: 1211

76 AlaAaGlyArgLeuAspArg 82
|||||
507 GCTGCCGTCGCTTGACCGC 487

seq_name: /cgn2.6/ptodata/1/lna/5D_COMB.seq:US-08-997-362-40

seq_documentation_block:
; Sequence 40, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-997-362-40
;
; alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-252-691-7056 x US-08-997-362-40/rev ..
;
; Align seg 1/1 to reverse of: US-08-997-362-40 from: 1 to: 1211
;
; 76 AlaAlaGlyArgLeuAspArg 82
; |||||
; 507 GCTGCCGTCGCGTTGACCGC 487
;
; seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-873-970-40
;
; seq_documentation_block:
; Sequence 40, Application 05/08873970
; Patent No. 6001361
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,970
; FILING DATE:
;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002C1
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-873-970-40
;
; alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-252-691-7056 x US-08-873-970-40/rev ..
;
; Align seg 1/1 to reverse of: US-08-873-970-40 from: 1 to: 1211
;
; 76 AlaAlaGlyArgLeuAspArg 82
; |||||
; 507 GCTGCCGTCGCGTTGACCGC 487
;
; seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-468-709B-5
;
; seq_documentation_block:
; Sequence 5, Application US/08468709B
; Patent No. 5654137
; GENERAL INFORMATION:
; APPLICANT: Astrom, Anders
; APPLICANT: Voorhees, John
; APPLICANT: Patterson, Ulrika
; APPLICANT: Tavakkol, Amir
; TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: PO Box 828
; CITY: Bloomfield Hills
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,709B
; FILING DATE: 06/06/95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36,683
; REFERENCE/DOCKET NUMBER: 2115-00676DVF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; TELEX: 287637
; INFORMATION FOR SEQ ID NO: 5:

```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 1322 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Placenta
IMMEDIATE SOURCE:
LIBRARY: human placenta genomic library
CLONE: lambda 2.1
FEATURE:
NAME/KEY: TATA_signal
LOCATION: 1008..1013
FEATURE:
NAME/KEY: exon
LOCATION: 1039..1245
FEATURE:
NAME/KEY: intron
LOCATION: 1246..1322
PUBLICATION INFORMATION:
AUTHORS: Astrom, Anders
AUTHORS: Pettersson, Ulrika
AUTHORS: Voorhees, John J
TITLE: Structure of the human cellular retinoic
TITLE: acid-binding protein II (CRABP-II) gene: Early
TITLE: transcriptional regulation by retinoic acid
JOURNAL: J. Biol. Chem.
DATE: 1992
RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 1322
US-08-468-709B-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-468-709B-5/rev ..
Align seg 1/1 to reverse of: US-08-468-709B-5 from: 1 to: 1322

27 SerSerArgGlnAlaThrArg 33
|||||
399 AGTAGTCGCCAGCGACTCGC 379

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-241-664B-5
seq_documentation_block:
Sequence 5, Application US/08241664B
Patent No. 5871909
GENERAL INFORMATION:
APPLICANT: Voorhees, John J.
APPLICANT: Astrom, Anders
APPLICANT: Pettersson, Ulrika
APPLICANT: Tavakkol, Amir
TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: PO Box 828
CITY: Bloomfield Hills
STATE: Michigan
COUNTRY: United States of America
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/241.664B
FILING DATE: May 11, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36,683
REFERENCE/DOCKET NUMBER: 2115-00676COD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
TELEX: 287637
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1322 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Placenta
IMMEDIATE SOURCE:
LIBRARY: human placenta genomic library
CLONE: lambda 2.1
FEATURE:
NAME/KEY: TATA_signal
LOCATION: 1008..1013
FEATURE:
NAME/KEY: exon
LOCATION: 1039..1245
FEATURE:
NAME/KEY: intron
LOCATION: 1246..1322
PUBLICATION INFORMATION:
AUTHORS: Astrom, Anders
AUTHORS: Pettersson, Ulrika
AUTHORS: Voorhees, John J
TITLE: Structure of the human cellular retinoic
TITLE: acid-binding protein II (CRABP-II) gene: Early
TITLE: transcriptional regulation by retinoic acid
JOURNAL: J. Biol. Chem.
DATE: 1992
RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 1322
US-08-241-664B-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-241-664B-5/rev ..
Align seg 1/1 to reverse of: US-08-241-664B-5 from: 1 to: 1322

27 SerSerArgGlnAlaThrArg 33
|||||
399 AGTAGTCGCCAGCGACTCGC 379

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03936-5
seq_documentation_block:
Sequence 5, Application PC/TUS9303936
GENERAL INFORMATION:
APPLICANT: Voorhees, John J.
APPLICANT: Astrom, Anders
APPLICANT: Pettersson, Ulrika
APPLICANT: Tavakkol, Amir
```

;; TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Harness, Dickey & Pierce
;; STREET: PO Box 828
;; CITY: Bloomfield Hills
;; STATE: Michigan
;; COUNTRY: United States of America
;; ZIP: 48013
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/03936
;; FILING DATE: 19930427
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Lewak, Anna M.
;; REGISTRATION NUMBER: 33,006
;; REFERENCE/DOCKET NUMBER: 211500676POB
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (313) 641-1600
;; TELEFAX: (313) 641-0270
;; TELEX: 287637
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1322 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; TISSUE TYPE: Placenta
;; IMMEDIATE SOURCE:
;; LIBRARY: human placenta genomic library
;; CLONE: lambda 2.1
;; FEATURE:
;; NAME/KEY: TATA_signal
;; LOCATION: 1008..1013
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 1039..1245
;; FEATURE:
;; NAME/KEY: intron
;; LOCATION: 1246..1322
;; PUBLICATION INFORMATION:
;; AUTHORS: Astrom, Anders
;; AUTHORS: Pettersson, Ulrika
;; AUTHORS: Voorhees, John J
;; TITLE: Structure of the human cellular retinolic
;; TITLE: acid-binding protein II (CRABP-II) gene: Early
;; TITLE: transcriptional regulation by retinoic acid
;; JOURNAL: J. Biol. Chem.
;; DATE: 1992
;; RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 1322
PCT-US93-03936-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x PCT-US93-03936-5/rev ..

Align seg 1/1 to reverse of: PCT-US93-03936-5 from: 1 to: 1322

27 SerSerArgGlnAlaThrArg 33
|||||
399 AGTAGTCGCAGCGACTCGC 379
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-389-812-3
seq_documentation_block:
; Sequence 3, Application US/08389812
; Patent No. 6015665
; GENERAL INFORMATION:
; APPLICANT: DIXIT, VISHVA M.
; TITLE OF INVENTION: METHOD AND COMPOSITION FOR REGULATING
; TITLE OF INVENTION: APOPTOSIS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,812
; FILING DATE: 13-FEB-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: KOSKI, ANTOINETTE F.
; REGISTRATION NUMBER: 34,202
; REFERENCE/DOCKET NUMBER: 20344-21036.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1468 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 295..1317
; US-08-389-812-3

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-389-812-3 ..

Align seg 1/1 to: US-08-389-812-3 from: 1 to: 1468

127 LeuArgAsnGlyValThrLeu 133
|||||
950 TTAAGAAATGCTGTGACTCTA 970

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-583-318-4

seq_documentation_block:
; Sequence 4, Application US/08583318
; Patent No. 5693483
; GENERAL INFORMATION:
; APPLICANT: Staunton, Donald
; APPLICANT: Harris, Edith
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin

;; TITLE OF INVENTION: Binding
;; NUMBER OF SEQUENCES: 14
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 233 South Wacker Drive, 6300 Sears Tower
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/583,318
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Williams Jr., Joseph A.
;; REGISTRATION NUMBER: 38,659
;; REFERENCE/DOCKET NUMBER: 27866/33050
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312-474-6300
;; TELEFAX: 312-474-0448
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1631 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..1437
;; US-08-583-318-4

alignment_scores: Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x US-08-583-318-4/rev ..

Align seg 1/1 to reverse of: US-08-583-318-4 from: 1 to: 1631

124 LeuAlaLysLeuArgAsnGly 130
|||||
488 TTGCAAACTGCGGAATGCG 468

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

OM of: US-09-252-691-7056 to: EST:* out_format : pfs

Date: Jun 10, 2000 2:23 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+2n.model -DEV=xlp  
-Q/cgn2-1/USPTO.spool/US09252691/runat_05062000_101737_1252/app_query.fasta.1  
-DB-EST -QWMT-fastap -SUFFIX=oligo.rst -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000  
-GAPEXT=0.050 -GAPEXT=0.050 -GAPEXT=60.000 -GAPEXT=60.000  
-GAPEXT=60.000 -GAPEXT=7.000 -GAPEXT=60.000 -GAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=1000 -DOCLIGN=200 -THR_SCORE=quality  
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=1000000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT  
-THREADS=1
```

Search information block:

Query: US-09-252-691-7056

Query length: 222

Database: EST*

Database sequences: 4857316

Database length: 202611650

Search time (sec): 604.990000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=60.000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb_est26:AI373988	+	52.00	990.63	283	AI373988 SMOVAFCAP32B02SK Oncho
em_est20:AW017830	+	25.00	467.99	295	AW017830 G1406A09.xl G14 - roo
gb_gss11:AA212974	+	10.00	173.41	0.7289	AQ212974 HS_3214_B2_E03_MR CIT
gb_est11:AA237292	-	8.00	143.47	33.90	AA237292 mw56d12.rl Soares mous
gb_est19:AA55409	+	8.00	140.82	47.63	AA55409 v53e06.rl Knowles Sol
gb_est39:AV206648	+	8.00	140.35	50.62	AV206648 AV206648 RIKEN full-le
gb_est12:AA286664	+	8.00	140.32	50.81	AA286664 vb79h04.rl Soares mous
gb_est18:AA688759	+	8.00	140.12	52.11	AA688759 vr07b09.rl Knowles Sol
gb_est1:W37395	-	8.00	139.71	54.90	X73795 CHESTM092 Goat mammary
gb_est27:AI424418	+	8.00	139.21	58.59	AI424418 tes0f02.rl NCI-CGAP_P
gb_est42:AV309318	+	8.00	138.89	60.98	AV309318 AV309318 RIKEN full-le
gb_est22:AI047546	+	8.00	138.14	67.19	AI047546 uh80b05.rl Soares mous
gb_est26:AI329557	+	8.00	137.89	69.37	AI329557 b6c08ne.rl Neurospora
gb_est13:AW047543	+	8.00	137.75	70.64	AW047543 UI-M-BH1-ama-f-02-0-UI
gb_est1:TS2062	+	8.00	137.42	73.71	TS2062 vb29d04.rl Stratagene fe
gb_est9:AA029259	-	8.00	137.05	77.31	AA029259 zn07b11.rl Stratagene
gb_est9:AA069085	-	8.00	136.97	78.02	AA069085 zm10c07.rl Stratagene
gb_est1:ID27082	+	8.00	136.88	78.92	ID27082 CELK0086CF Yuji Kohara
gb_est9:AA089014	+	8.00	136.85	79.28	AA089014 mn85e11.rl Stratagene
gb_est39:AL004306	+	8.00	136.62	81.60	AL004306 DRZp434M1013.rl 434
gb_est39:AW123360	+	8.00	136.57	82.14	AW123360 UI-M-BH2-1-apf-e-10-0-
gb_est27:AI398322	+	8.00	136.51	82.85	AI398322 NCSM289T3 Subtracted M
gb_est25:AI273549	+	8.00	136.26	85.52	AI273549 q155d03.rl Soares mous
gb_est6:W09815	-	8.00	136.13	86.94	W09815 ma49e07.rl Soares mous
gb_est6:W09817	-	8.00	136.07	87.65	W09817 ma49e07.rl Soares mous
gb_est6:W09817	-	8.00	136.03	88.01	W09817 ma49f01.rl Soares mous
gb_est7:W030249	+	8.00	135.91	89.42	W30249 mc28b01.rl Soares mous
gb_est23:AI140187	+	8.00	135.79	90.84	AI140187 qe20b09.rl Soares_feta
gb_est19:AA293606	-	8.00	135.56	93.49	AA293606 vx90e08.rl Stratagene
gb_est11:AA250690	-	8.00	135.49	94.37	AA250690 mp79h02.rl Soares mous
gb_est9:AA072069	-	8.00	135.42	95.25	AA072069 mm59g06.rl Stratagene
gb_est11:AA270914	+	8.00	135.42	95.25	AA270914 va70c06.rl Soares mous
gb_est9:AA060088	-	8.00	135.29	96.83	AA060088 mj71b09.rl Soares mous
gb_est18:AA674188	+	8.00	135.16	98.41	AA674188 vp97c02.rl Stratagene
gb_est26:AI329555	+	8.00	135.14	98.76	AI329555 b6b12he.rl Neurospora
gb_gss5:AW078905	+	8.00	135.14	98.76	AW078905 HS_3208_B2_D04_MR CIT
gb_est37:AW012406	+	8.00	135.08	99.46	AW012406 uo04e08.rl Sugano mous
gb_gss13:AA367068	+	8.00	135.04	99.99	AA367068 HS_5032_B1_F05_T7 RPT
gb_est9:AA086716	-	8.00	134.89	101.91	AA086716 mn93h11.rl Stratagene

gb_est15:AA516785	-	8.00	134.89	101.91	AA516785 vh88d10.rl Knowles
gb_est11:AA268935	-	8.00	134.88	102.09	AA268935 va99g10.rl Soares m
gb_est11:AA266576	-	8.00	134.86	102.26	AA266576 m25e10.rl Barstead
gb_est44:AW211779	-	8.00	134.86	102.26	AW211779 uc85e01.rl NCI-CGAP
gb_est34:AI786062	-	8.00	134.73	104.01	AI786062 u157h02.yl Sugano m
gb_est15:AA75027	-	8.00	134.64	105.23	AA75027 v403f09.rl Soares_m
gb_est24:AI182222	+	8.00	134.64	105.23	AI182222 uc68c04.rl Soares_m
gb_est14:AA45266	-	8.00	134.63	105.41	AA45266 vf62h11.rl Barstead
gb_est37:AI954969	-	8.00	134.45	107.84	AI954969 wx92b02.rl NCI-CGAP
gb_est35:AI853843	-	8.00	134.41	108.37	AI853843 UI-M-BH0-a1j-r-d-02-0
gb_gss7:QA0972312	+	8.00	134.41	108.37	QA0972312 RPCI-23-312L19.TV R
gb_gss12:AI035497	-	8.00	134.36	109.06	AI035497 ub48b04.rl Soares_m
gb_gss12:QA273411	+	8.00	134.26	110.45	QA273411 nbx003022rlr CUGI R
gb_gss9:AA025734	-	8.00	134.25	110.62	AA025734 l1(2)k0403 Drosophi
gb_gss7:QA948077	+	8.00	134.14	112.19	QA948077 Sheared DNA-50NT.TF
gb_gss7:QA9363138	+	8.00	134.09	112.88	QA9363138 nbx0055A20f CUGI R
gb_est20:AA529215	-	8.00	134.01	114.09	AA529215 u161c06.yl Sugano m
gb_est20:AA869364	-	8.00	133.89	115.82	AA869364 vq07b11.rl Barstead
gb_est25:AI303127	-	8.00	133.57	120.65	AI303127 u160b05.yl Sugano m
gb_est24:AI226044	-	8.00	133.37	123.52	AI226044 u108f03.yl Sugano m
gb_gss1:AG010516	-	8.00	132.73	134.52	AG010516 Homo sapiens genom
gb_gss1:AG010526	-	8.00	132.64	136.05	AG010526 Homo sapiens genom
gb_est29:AI573796	-	8.00	132.15	144.88	AI573796 u165h03.yl Sugano m
gb_est23:AI119391	+	8.00	131.75	152.48	AI119391 uf03f07.yl Sugano m
gb_gss3:BL1311	+	8.00	131.18	164.06	BL1311 F22P19-Sp6 IGF Arabid
gb_est22:AI040945	+	7.00	131.36	160.24	AI040945 ov53b11.xl Soares_te
gb_est37:AI992098	-	7.00	126.99	280.56	AI992098 w21g10.xl NCI-CGAP
gb_est11:TI0501	-	7.00	126.06	316.32	TI0501 hbc553 Human pancreat
gb_est29:AI612921	-	7.00	125.72	330.52	AI612921 tx16g01.xl NCI-CGAP
gb_est33:AI096038	-	7.00	125.39	344.66	AI096038 SMOVL3CAN20E03 Onch
gb_est33:AV087819	-	7.00	125.28	349.37	AV087819 AV087819 Mus muscul
gb_est38:AW031720	+	7.00	125.18	354.06	AW031720 EST275174 tomato ca
gb_est44:AA76501	-	7.00	123.74	426.16	AA76501 RC4-CR0036-180899-0
gb_est5:U38048	+	7.00	123.65	430.77	U38048 OSU38048 FRSC Oryza
gb_gss13:AA0420961	-	7.00	122.94	472.06	AA0420961 RPCI-11-185L24.TJ R
gb_est18:AA727187	+	7.00	122.86	476.63	AA727187 vu91c11.rl Stratage
gb_est24:AI218764	-	7.00	122.79	481.19	AI218764 SMOVAFCAP12B07SK On
gb_est46:AW315321	-	7.00	122.79	481.19	AW315321 UI-R-B70-adv-b-09-0
gb_est17:AA028260	-	7.00	122.36	508.48	AA028260 SMOVL3CA6555 Oncho
gb_gss7:QA928279	-	7.00	122.32	510.75	QA928279 RPCI-23-262D8.TJ RP
gb_est29:AI625663	-	7.00	122.25	515.29	AI625663 ty58g05.xl NCI-CGAP
gb_est16:CA29055	-	7.00	122.15	522.08	CA29055 C29055 Rice callus CD
gb_gss14:AA67486	+	7.00	122.12	524.34	AA67486 HS_5219.A2-C05-SP6E
gb_est34:AV150321	-	7.00	121.95	535.64	AV150321 AV150321 Mus muscul
gb_est7:W33989	-	7.00	121.85	542.41	W33989 mb57e04.rl Soares mou
gb_est43:AW176774	+	7.00	121.79	546.92	AW176774 RCO-CT0094-280899-0
gb_est43:AV346666	+	7.00	121.76	549.17	AV346666 AV346666 RIKEN full
gb_est3:R73420	+	7.00	121.69	553.67	R73420 YJ92h10.rl Soares bre
gb_est44:AV377646	-	7.00	121.54	564.92	AV377646 AV377646 RIKEN full
gb_est34:AI790077	-	7.00	121.48	569.41	AI790077 ue67c08.rl Soares_m
gb_est46:AW312649	-	7.00	121.48	569.41	AW312649 4784 MARC 1P1G Sus
gb_est42:AV287139	+	7.00	121.45	571.65	AV287139 AV287139 RIKEN full
gb_est13:AA362027	+	7.00	121.38	576.14	AA362027 EST71371.T-cell lym
gb_est23:AA253941	-	7.00	121.35	578.38	AA253941 aq28f07.xl Barstead
gb_est12:AA299011	+	7.00	121.32	580.62	AA299011 EST11467 Uterus Hom
gb_est24:AI2139	+	7.00	121.26	585.10	AI2139 HSC04A05.xl NCI-CGAP
gb_est21:242139	+	7.00	121.23	587.34	242139 HSC04A05 normalized
gb_est28:AI503961	-	7.00	121.23	587.34	AI503961 vm40d06.xl Knowles
gb_est43:AV362827	+	7.00	121.20	589.58	AV362827 AV362827 RIKEN full
gb_est43:AV341813	+	7.00	121.03	602.99	AV341813 AV341813 RIKEN full
gb_est8:AA015004	-	7.00	121.00	605.22	AA015004 mm26a03.rl Soares m
gb_est34:AI173917	-	7.00	121.00	605.22	AI173917 AV173917 Mus muscul
gb_est16:CA23540	-	7.00	120.91	611.91	CA23540 C23540 Japanese floun
gb_gss10:AA701318	-	7.00	120.89	614.14	AA701318 HS_3102.A2-G10_MR C
gb_est47:AA437009	-	7.00	120.86	616.37	AA437009 77688 MARC 12F1G.Sus
gb_est21:AA380755	-	7.00	120.75	625.28	AA380755 ua45c05.s1 Soares_m
gb_est46:AA350069	-	7.00	120.75	625.28	AA350069 GM210006B20F3R Gm-r
gb_est47:AA431479	-	7.00	120.75	625.28	AA431479 72057 MARC 2P1G Sus
gb_est41:AV268759	-	7.00	120.66	631.95	AV268759 AV268759 RIKEN full
gb_est10:AA179869	-	7.00	120.58	638.62	AA179869 zp38h01.rl Stratage
gb_est1:240954	-	7.00	120.55	640.84	240954 HSC2KB092 normalised

gb_est17:C65099	7.00	118.07	881.24	374	C65099	C65099	Yuj1 Kohara unpub	gb_gss12:AQ293177	7.00	117.38	962.79	412	AQ293177	HS_2201_B2_H04_MF_C
gb_est25:AI309404	7.00	118.05	883.40	375	AI309404	tb37c09.x1 NCI CGAP HS	gb_gss13:AQ422850	7.00	117.38	962.79	412	AQ422850	CITB1-E1-2577G12.TF	
gb_est8:AA039025	7.00	118.03	885.55	376	AA039025	m199e04.x1 Soares mous	gb_est15:AA479663	7.00	117.35	967.07	414	AA479663	qv18h01.x1 Soares_N	
gb_est16:CA44658	7.00	118.03	885.55	376	CA44658	C44658 Yuj1 Kohara unpub	gb_est26:AI361973	7.00	117.35	967.07	414	AI361973	qv48e11.x1 NCI CGAP	
gb_est17:C71614	7.00	118.03	885.55	376	C71614	C71614 Yuj1 Kohara unpub	gb_est45:AW244976	7.00	117.31	969.20	415	AW244976	687017H02.y1 687 -	
gb_est19:AA747504	7.00	118.03	885.55	376	AA747504	ay77c03.x1 NCI CGAP_Ew	gb_est22:AI074659	7.00	117.31	971.34	416	AI074659	ex82g07.x1 Soares_N	
gb_gss10:AQ09298	7.00	118.01	887.71	377	AQ09298	HS_3042_B2_A01_MF_CIT	gb_est15:AA494356	7.00	117.29	973.48	417	AA494356	ne38h08.s1 NCI CGAP	
gb_est1:3D4021	7.00	117.99	889.86	378	3D4021	CELK039GZr Yuj1 Kohara u	gb_est22:AI003949	7.00	117.29	973.48	417	AI003949	ot34b08.x1 Soares_N	
gb_est24:AI166137	7.00	117.99	889.86	378	AI166137	B008p30u Hybrid aspen	gb_est22:AI155914	7.00	117.29	973.48	417	AI155914	ue04a08.x1 Soares_N	
gb_est29:AI566957	7.00	117.99	889.86	378	AI566957	tn25c09.x1 NCI CGAP_Br	gb_est14:AA398211	7.00	117.28	975.61	418	AA398211	tm59c05.x1 Soares_N	
gb_est36:AV201161	7.00	117.99	889.86	378	AV201161	AV201161 Yuj1 Kohara u	gb_est28:AI502971	7.00	117.28	975.61	418	AI502971	vm69c05.x1 Knowles	
gb_est28:AI490820	7.00	117.94	896.32	381	AI490820	EST241529 Tomato shoot	gb_est15:AA456702	7.00	117.26	977.74	419	AA456702	au13f11.r1 Soares_N	
gb_gss15:AQ522581	7.00	117.94	896.32	381	AQ522581	HS_5240_A2_F05_77A_RPC	gb_est4:HS1345	7.00	117.26	977.74	419	HS1345	yo30405.r1 Soares_N	
gb_est4:062400	7.00	117.92	898.47	382	062400	H0M272A05B Clontech huma	gb_est35:AI820692	7.00	117.24	979.88	420	AI820692	yj87c12.y5 Soares_b	
gb_est18:AA666029	7.00	117.92	898.47	382	AA666029	ag44b12.s1 Jia Bone ma	gb_gss16:AAQ30326	7.00	117.24	979.88	420	AAQ30326	RPCI-11-469J24.TJ R	
gb_est31:AI852708	7.00	117.92	898.47	382	AI852708	UI-M-BH0-aj1-e11-0-UI	gb_est21:AAQ35182	7.00	117.23	982.01	421	AAQ35182	ot33d07.s1 NCI CGAP	
gb_gss9:AOQ31405	7.00	117.92	898.47	382	AOQ31405	HS_2222_A2_E07_MF_CIT	gb_est24:AA020915	7.00	117.23	982.01	421	AA020915	AU029915 Rice cdNA	
gb_est6:W11319	7.00	117.90	900.62	383	W11319	ma78c05.r1 Soares mous	gb_est26:AA008984	7.00	117.23	982.01	421	AA008984	UI-R-EO-dc-a-10-0-U	
gb_est16:AA561354	7.00	117.86	904.92	385	AA561354	vl18e08.r1 Stratagene	gb_est29:AI566161	7.00	117.23	982.01	421	AI566161	tg69b11.x1 NCI CGAP	
gb_est8:CI0522	7.00	117.85	907.07	386	CI0522	C10522 Yuj1 Kohara unpub	gb_est32:AI820692	7.00	117.23	982.01	421	AI820692	wh97a08.x1 NCI CGAP	
gb_est9:AA107760	7.00	117.85	907.07	386	AA107760	mo48g03.r1 Life Tech m	gb_est45:AW258986	7.00	117.23	982.01	421	AW258986	um76a04.y1 Sugano m	
gb_est29:AI569458	7.00	117.85	907.07	386	AI569458	tn87c02.x1 NCI CGAP_Ut	gb_gss11:AAQ119894	7.00	117.23	982.01	421	AAQ119894	HS_3252_A2_G11_MR_C	
gb_est37:AI973249	7.00	117.83	909.22	387	AI973249	tn53f12.x1 NCI CGAP_Ut	gb_est30:AI662710	7.00	117.21	984.15	422	AI662710	va86h11.x1 Soares m	
gb_est44:AW214739	7.00	117.83	909.22	387	AW214739	uo99c09.y1 NCI CGAP_Lu	gb_gss5:AOQ07397	7.00	117.21	984.15	422	AOQ07397	HS_3177_A2_D11_MR_C	
gb_est15:AA535707	7.00	117.79	913.52	389	AA535707	nf88d04.s1 NCI CGAP_Cc	gb_est25:AI307402	7.00	117.19	986.28	423	AI307402	tb26h10.x1 NCI CGAP	
gb_est19:AA688437	7.00	117.79	913.52	389	AA688437	ot22f07.s1 NCI CGAP_Ki	gb_gss13:AAQ371431	7.00	117.18	988.41	424	AAQ371431	HS_5043_A1_F04_SP6E	
gb_est32:AI715296	7.00	117.79	913.52	389	AI715296	UI-R-Y-0-abb-f-12-0-UI	gb_gss14:AAQ402925	7.00	117.18	988.41	424	AAQ402925	HS_5066_A2_E12_SP6E	
gb_gss3:R33820	7.00	117.79	913.52	389	R33820	HS-1023-A2-E04-MF.abi CI	gb_gss14:AAQ499762	7.00	117.18	988.41	424	AAQ49762	HS_5204_A2_B09_SP6E	
gb_gss3:R42625	7.00	117.75	913.52	389	R42625	HS-1056-A1-G08-MF.abi CI	gb_est20:AAQ070578	7.00	117.16	990.54	425	AAQ070578	vq23d10.r1 Barstead	
gb_gss12:AQ312357	7.00	117.75	913.52	391	AQ312357	RPCI11-117M18.TJ RPCI	gb_est37:AW008196	7.00	117.14	992.67	426	AW008196	wv51d02.x1 NCI CGAP	
gb_est11:AA222474	7.00	117.74	919.96	392	AA222474	mw20e02.r1 Soares mous	gb_gss4:AOQ675929	7.00	117.14	992.67	426	AOQ675929	HS_2159_A1_G10_WTC	
gb_est39:AA123302	7.00	117.74	919.96	392	AA123302	UI-M-BH2-1-agg-h-11-0-	gb_gss5:AAQ032091	7.00	117.14	992.67	426	AAQ032091	HS_5243_B1_E11_SP6E	
gb_est3:R65616	7.00	117.72	922.10	393	R65616	Y126b00.r1 Soares placen	gb_gss13:AAQ43791	7.00	117.13	994.80	427	AAQ43791	HS_5049_A1_C01_SP6E	
gb_est44:AW206298	7.00	117.70	924.25	394	AW206298	UI-H-B11-278D3-4.TV RPCI	gb_est37:AAQ436992	7.00	117.11	996.94	428	AAQ43692	E5T01998 Mouse 7.5	
gb_gss14:AO507837	7.00	117.70	924.25	394	AO507837	RPCI-11-278D3-4.TV RPCI	gb_est14:AAQ406957	7.00	117.08	1.0e+03	430	AAQ40692	E5T01998 Mouse 7.5	
gb_est28:W89921	7.00	117.68	926.39	395	W89921	mf77c09.r1 Soares mous	gb_est16:AA583432	7.00	117.08	1.0e+03	430	AA583432	tn38a08.s1 NCI CGAP	
gb_est29:AA566323	7.00	117.66	928.54	396	AA566323	ve24b03.y1 Soares mous	gb_est20:AAQ36857	7.00	117.06	1.0e+03	430	AAQ36857	of35a05.s1 NCI CGAP	
gb_est34:AI801623	7.00	117.66	928.54	396	AI801623	ts01h03.x1 NCI CGAP_Ga	gb_est6:79372	7.00	117.06	1.0e+03	431	79372	HUM281G04B Human aort	
gb_gss11:R335674	7.00	117.66	928.54	396	R335674	HS-1029-A2-H07-MR.abi CI	gb_gss11:AAQ08520	7.00	117.06	1.0e+03	431	AAQ08520	HS_3237_A1_A04_MR_C	
gb_gss11:AI977979	7.00	117.65	930.68	397	AI977979	HS_3142-B2_E07_MR.abi CI	gb_est7:W44184	7.00	117.04	1.0e+03	432	W44184	mc69e07.r1 Soares mou	
gb_est21:AA423456	7.00	117.63	932.83	398	AA423456	tf725h04.s1 NCI CGAP_Bi	gb_est18:AA694431	7.00	117.04	1.0e+03	432	AA694431	av22b05.s1 Soares_P	
gb_est21:AA989124	7.00	117.61	934.97	399	AA989124	ot75h04.s1 NCI CGAP_Lu	gb_est23:AI082401	7.00	117.04	1.0e+03	432	AI082401	ov22c10.s1 NCI CGAP	
gb_est11:R00533	7.00	117.59	937.11	400	R00533	WEST01254 Early embryo,	gb_est27:AI421702	7.00	117.04	1.0e+03	432	AI421702	tf54d09.x1 NCI CGAP	
gb_est15:AA515027	7.00	117.57	939.26	401	AA515027	nf64h06.s1 NCI CGAP_Cc	gb_est39:AW126197	7.00	117.04	1.0e+03	432	AW126197	um20c12.y1 Sugano m	
gb_gss7:AOQ03740	7.00	117.57	939.26	401	AOQ03740	G58TC04781 Trypanosome	gb_est39:AW126197	7.00	117.04	1.0e+03	432	AW126197	um20c12.y1 Sugano m	
gb_gss10:AOQ101952	7.00	117.57	939.26	401	AOQ101952	HS_3017_B1_G12-T7 CIT	gb_gss4:AOQ704266	7.00	117.04	1.0e+03	432	AOQ704266	HS_5475_B2_H10_SP6E	
gb_gss11:AO186130	7.00	117.57	939.26	401	AO186130	HS_3077_A1_B11_MR_CIT	gb_est19:AA760813	7.00	117.03	1.0e+03	433	AA760813	nz10f05.s1 NCI CGAP	
gb_est2:R97145	7.00	117.56	941.40	402	R97145	ve50a10.r1 Soares fetal	gb_est19:AA760813	7.00	117.03	1.0e+03	433	AA760813	nz10f05.s1 NCI CGAP	
gb_gss3:R67144	7.00	117.56	941.40	402	R67144	CIT-HSP-2017024.TFB CIT-	gb_est20:AA822155	7.00	117.03	1.0e+03	433	AA822155	vp35e08.r1 Barstead	
gb_est42:AA162091	7.00	117.54	943.54	403	AA162091	au89d07.y1 Schneider f	gb_est28:AI506159	7.00	117.03	1.0e+03	433	AI506159	vp35e08.r1 Barstead	
gb_est3:R24986	7.00	117.52	945.68	404	R24986	Yg41h03.r1 Soares infant	gb_est5:R26282	7.00	116.99	1.0e+03	435	R26282	Yw62b10.s1 Soares_Pla	
gb_est24:AI424291	7.00	117.52	945.68	404	AI424291	q116g02.x1 Soares NHM	gb_est10:AA157978	7.00	116.99	1.0e+03	435	AA157978	zo71h07.s1 Stratage	
gb_est17:AA634492	7.00	117.50	947.82	405	AA634492	zu76d01.s1 Soares test	gb_est10:AA162230	7.00	116.99	1.0e+03	435	AA162230	mn44h01.r1 Bedding	
gb_est34:AI795818	7.00	117.50	947.82	405	AI795818	RCO-HT0023-300699-001-	gb_est22:AI004733	7.00	116.99	1.0e+03	435	AI004733	ou04e06.x1 Soares_t	
gb_est16:C29077	7.00	117.49	949.96	406	C29077	C39077 Rice callus cdNA	gb_est29:C95514	7.00	116.99	1.0e+03	435	C95514	C95514 Citrus unshu	
gb_est38:AA078385	7.00	117.49	949.96	406	AA078385	AU078385 Rice green sh	gb_est32:AI762505	7.00	116.98	1.0e+03	436	AI762505	wn91d02.x1 NCI CGAP	
gb_gss16:AO620418	7.00	117.47	952.10	407	AO620418	HS_5180_B1_B07_SP6E RH	gb_gss4:AO668655	7.00	116.96	1.0e+03	437	AO668655	nbx007301.s1 CUGI R	
gb_est15:AA458096	7.00	117.45	954.24	408	AA458096	zw49g08.r1 Soares tota	gb_gss9:R95967	7.00	116.96	1.0e+03	437	R95967	T30K5TR TAMU Arabidop	
gb_est15:AA457407	7.00	117.45	954.24	408	AA457407	aa86f03.s1 Stratagene	gb_est4:R57179	7.00	116.93	1.0e+03	439	R57179	HUM274D04B Clontech h	
gb_est47:AAW435314	7.00	117.45	954.24	408	AAW435314	fa12a07.s1 Zebrafish m	gb_est42:AW161562	7.00	116.93	1.0e+03	439	AW161562	au82h03.y1 Schneide	
gb_est4:DG0684	7.00	117.43	956.38	409	DG0684	HUM125D12A Clontech huma	gb_est25:AI278489	7.00	116.91	1.0e+03	440	AI278489	gm34h08.x1 NCI CGAP	
gb_gss1:AOQ47317	7.00	117.43	956.38	409	AOQ47317	RPCI11-32014.TK RPCI-11	gb_est28:AI524059	7.00	116.91	1.0e+03	440	AI524059	th01c03.x1 NCI CGAP	
gb_gss9:AIQ177132	7.00	117.43	956.38	409	AIQ177132	HS_3202_A2_A08_MR_CIT	gb_est21:AA942796	7.00	116.90	1.0e+03	441	AA942796	EST198295 Normalize	
gb_est28:AI466197	7.00	117.40	960.66	411	AI466197	vr82e09.y1 Stratagene	gb_est25:AI245725	7.00	116.90	1.0e+03	441	AI245725	ek36e03.x1 NCI CGAP	
gb_est31:AI672708	7.00	117.40	960.66	411	AI672708	we57f10.x1 Soares_thym	gb_est8:AA009744	7.00	116.88	1.0e+03	442	AA009744	zg82g01.s1 Soares_f	
gb_est46:AAW32288	7.00	117.40	960.66	411	AAW32288	cm2-HT0326-291199-044-	gb_gss11:AAQ35415	7.00	116.88	1.0e+03	442	AAQ35415	mexb0012P10r CUGI R	
gb_gss3:R95591	7.00	117.40	960.66	411	R95591	RPCI11-20C11.TV RPCI-11	gb_est11:AAQ154861	7.00	116.86	1.0e+03	443	AA154861	mq61f03.r1 Soares_t	
gb_est20:AAQ897470	7.00	117.38	962.79	412	AAQ897470	al48e04.s1 Soares_NFL	gb_est24:AI190323	7.00	116.86	1.0e+03	443	AI190323	qd57h06.x1 Soares_t	
gb_est35:AA044310	7.00	117.38	962.79	412	AA044310	DKF293430A0429.r1 434 C	gb_gss9:AIQ190323	7.00	116.86	1.0e+03	443	AIQ190323	qd57h06.x1 Soares_t	
gb_est45:AW245695	7.00	117.38	962.79	412	AW245695	R282917.5Prime NTH_MGC	gb_est2:R08275	7.00	116.85	1.0e+03	444	R0827		

gb_ests3:BS2705	7.00	116.85	1.0e+03	444	B52705 CIT-HSP-200616.TR CIT-HSP	gb_ests23:AI147693	7.00	116.46	1.1e+03	469	AI147693 qb3a10.x1 NCI_CGAP
gb_ests9:AA11938	7.00	116.83	1.0e+03	445	AA11938 mp84d03.r1 Soares_thym	gb_ests26:AI347038	7.00	116.46	1.1e+03	469	AI347038 qb54h11.x1 NCI_CGAP
gb_ests10:AA135185	7.00	116.80	1.0e+03	447	AA135185 uo27a05.s1 Stratagene	gb_gss13:AI0101242	7.00	116.46	1.1e+03	469	AI0101242 HS_5066.A1.G01.SP6E
gb_ests11:AA175271	7.00	116.80	1.0e+03	447	AA175271 uo30b04.y1 Sugano mous	gb_ests20:F22874	7.00	116.44	1.1e+03	470	F22874 SSC15608 Porcine smal
gb_ests12:AA199418	7.00	116.80	1.0e+03	447	AA199418 U1555815-A.thaliana	gb_ests46:AAV0403482	7.00	116.44	1.1e+03	470	AAV0403482 AV403482 Bombyx mor
gb_ests13:AAW04639	7.00	116.80	1.0e+03	447	AAW04639 UI-W-BHI-akv-b-05-O-UT	gb_ests8:AAO36083	7.00	116.43	1.1e+03	471	AAO36083 mi73h02.r1 Soares m
gb_ests14:AAW073711	7.00	116.80	1.0e+03	447	AAW073711 xbo1h10.x1 NCI_CGAP_GU	gb_ests20:AA859032	7.00	116.43	1.1e+03	471	AA859032 vq25b02.r1 Barstead
gb_ests15:AAW160789	7.00	116.80	1.0e+03	447	AAW160789 au68b05.y1 Schneider f	gb_ests23:AI098344	7.00	116.43	1.1e+03	471	AI098344 vq84b12.r1 Barstead
gb_ests16:AAW291316	7.00	116.80	1.0e+03	447	AAW291316 UI-H-B12-agu-c-12-O-UT	gb_ests39:AAW130018	7.00	116.43	1.1e+03	471	AAW130018 vq26h03.x1 NCI_CGAP
gb_ests17:AAW60063	7.00	116.78	1.0e+03	448	AAW60063 vq69a03.r1 Stratagene	gb_ests7:W89543	7.00	116.41	1.1e+03	472	W89543 mf71c10.r1 Soares mou
gb_ests18:AAJ366778	7.00	116.78	1.0e+03	448	AAJ366778 qv41d06.x1 Soares_NHm	gb_ests27:AI432456	7.00	116.41	1.1e+03	472	AI432456 tg73g05.x1 Soares_N
gb_ests19:AAJ16334	7.00	116.78	1.0e+03	448	AAJ16334 au46c01.y1 Schneider f	gb_ests44:AI273531	7.00	116.41	1.1e+03	472	AI273531 AJ273531 Metarhiziu
gb_ests20:AAQ200590	7.00	116.77	1.0e+03	449	AAQ200590 RPCI11-43HF.TK RPCI-11	gb_ests46:AAQ18537	7.00	116.41	1.1e+03	472	AAQ18537 um99a02.y1 Sugano m
gb_ests21:AAQ18172	7.00	116.77	1.0e+03	449	AAQ18172 CEK0106YF.Yuji Kohara u	gb_ests31:AI695261	7.00	116.40	1.1e+03	473	AI695261 wa5e10.l1 Soares_N
gb_ests22:AAI19062	7.00	116.77	1.0e+03	449	AAI19062 q124f01.x1 NCI_CGAP_B4	gb_gss7:AAQ972473	7.00	116.40	1.1e+03	473	AAQ972473 RPCI-23-312021.TJ R
gb_ests23:AAI288600	7.00	116.77	1.0e+03	449	AAI288600 q182g12.x1 Soares_NHm	gb_ests15:AA528216	7.00	116.38	1.1e+03	474	AA528216 nj16h06.s1 NCI_CGAP
gb_ests24:AAQ931292	7.00	116.77	1.0e+03	449	AAQ931292 ul67c10.y1 Sugano mous	gb_gss5:AAQ811421	7.00	116.38	1.1e+03	474	AAQ811421 HS_5226.B1.D11.SP6E
gb_ests25:AAQ984707	7.00	116.77	1.0e+03	449	AAQ984707 RPCI-23-266015.TV RPCI	gb_ests7:AAQ941433	7.00	116.38	1.1e+03	474	AAQ941433 Sheared DNA-44M18.T
gb_ests26:AAQ122354	7.00	116.77	1.0e+03	449	AAQ122354 HS_3087.AI.D11.MR CIT	gb_ests8:AAQ98437	7.00	116.37	1.1e+03	475	AAQ98437 zbf77g09.s1 Soares_sen
gb_ests27:AAQ16114	7.00	116.75	1.0e+03	450	AAQ16114 ze31f08.r1 Soares retl	gb_ests24:AI2113328	7.00	116.37	1.1e+03	475	AI2113328 o6d01a1.r1 Aspergill
gb_ests28:AAQ18172	7.00	116.75	1.0e+03	450	AAQ18172 zv94f03.r1 Soares_NHm	gb_ests29:AI554501	7.00	116.37	1.1e+03	475	AI554501 tq22c09.x1 NCI_CGAP
gb_ests29:AAQ768438	7.00	116.75	1.0e+03	450	AAQ768438 OB22f08.s1 NCI_CGAP_K1	gb_ests7:W87316	7.00	116.35	1.1e+03	476	W87316 zh64d05.r1 Soares fet
gb_ests30:AAQ48553	7.00	116.75	1.0e+03	450	AAQ48553 mgxb0021A13f CUGI Rice	gb_ests17:C78791	7.00	116.35	1.1e+03	476	C78791 C78791 Mouse 3.5-dpc
gb_ests31:DA6560	7.00	116.74	1.0e+03	451	DA6560 RIGS11315A Rice green sh	gb_ests25:AI290167	7.00	116.35	1.1e+03	476	AI290167 q180d12.x1 Soares_N
gb_ests32:R25343	7.00	116.74	1.0e+03	451	R25343 yH42d05.r1 Soares placen	gb_gss13:AAQ389481	7.00	116.35	1.1e+03	476	AAQ389481 RPCI11-153H22.TV RP
gb_ests33:AAQ120600	7.00	116.74	1.0e+03	451	AAQ120600 mu71g01.r1 Soares mous	gb_gss13:AAQ389481	7.00	116.35	1.1e+03	476	AAQ389481 RPCI11-153H22.TV RP
gb_ests34:AAQ24494	7.00	116.74	1.0e+03	451	AAQ24494 EST27613 Cerebellum II	gb_gss14:AAQ316506	7.00	116.35	1.1e+03	476	AAQ316506 HS_5072.B2.A09.T7A
gb_ests35:AAQ684357	7.00	116.74	1.0e+03	451	AAQ684357 vq69a10.s1 Knowles Sol	gb_ests4:HA32375	7.00	116.34	1.1e+03	477	HA32375 ym57f04.s1 Soares inf
gb_ests36:AAQ29472	7.00	116.74	1.0e+03	451	AAQ29472 wx12b09.x1 NCI_CGAP_G4	gb_ests20:AA847787	7.00	116.32	1.1e+03	478	AA847787 oeg9h10.s1 NCI_CGAP
gb_ests37:AAQ05512	7.00	116.74	1.0e+03	451	AAQ05512 HS_3094.A2.H07.MR CIT	gb_ests38:AAW058087	7.00	116.32	1.1e+03	478	AAW058087 wv83b06.x1 Soares t
gb_ests38:AAQ1557599	7.00	116.72	1.0e+03	452	AAQ1557599 Pt2.1.1.G05.r tumor2 H	gb_ests5:W62295	7.00	116.31	1.1e+03	479	W62295 vq62f10.s1 Soares pla
gb_ests39:AAQ769083	7.00	116.72	1.0e+03	452	AAQ769083 HS_3155.A2.B11.MR CIT	gb_ests8:AAQ30424	7.00	116.31	1.1e+03	479	AAQ30424 ml28g04.r1 Soares m
gb_ests40:AAQ163527	7.00	116.70	1.0e+03	453	AAQ163527 ms22g12.r1 Stratagene	gb_ests23:AI091181	7.00	116.31	1.1e+03	479	AI091181 co18d07.x1 Sugano m
gb_ests41:AAQ263516	7.00	116.70	1.0e+03	453	AAQ263516 q120b03.x1 Soares_NHm	gb_ests45:AAW259944	7.00	116.31	1.1e+03	479	AAW259944 um79g01.x1 Sugano m
gb_ests42:AAQ1801537	7.00	116.70	1.0e+03	453	AAQ1801537 tq09f12.x1 NCI_CGAP_G4	gb_gss11:AAQ237059	7.00	116.31	1.1e+03	479	AAQ237059 HS_2024.B1.A10.MR C
gb_ests43:AAQ34841	7.00	116.69	1.0e+03	454	AAQ34841 m152f09.r1 Soares mous	gb_ests4:HA32375	7.00	116.28	1.1e+03	481	HA32375 ym57f04.s1 Soares inf
gb_ests44:AAQ35711	7.00	116.69	1.0e+03	454	AAQ35711 Homo sapiens mRNA; EST	gb_ests17:AA632915	7.00	116.28	1.1e+03	481	AA632915 nm14g01.s1 NCI_CGAP
gb_ests45:AAQ24834	7.00	116.69	1.0e+03	454	AAQ24834 Fugu rubripes GSS sequ	gb_ests27:AAQ389481	7.00	116.28	1.1e+03	481	AAQ389481 RPCI11-153H22.TV RP
gb_ests46:AAQ364491	7.00	116.67	1.0e+03	454	AAQ364491 nbx0061113r CUGI Rice	gb_ests45:AAW275342	7.00	116.28	1.1e+03	481	AAW275342 km79c07.x1 NCI_CGAP
gb_ests47:AAQ40342	7.00	116.67	1.0e+03	455	AAQ40342 EST283206 tomato mixed	gb_ests8:AAQ261675	7.00	116.26	1.1e+03	482	AAQ261675 nbh2g07.r1 Soares m
gb_ests48:AAQ02696	7.00	116.67	1.0e+03	455	AAQ02696 UI-HF-BK0-sav-g-06-O-UT	gb_gss7:AAQ364250	7.00	116.26	1.1e+03	482	AAQ364250 nbx0060P21f CUGI R
gb_ests49:AAQ69503	7.00	116.67	1.0e+03	455	AAQ69503 HS_5518.A2.H09.SP6E RH	gb_gss15:AAQ47651	7.00	116.26	1.1e+03	483	AAQ47651 RPCI-11-382C6.TJ RP
gb_ests50:AAQ75789	7.00	116.67	1.0e+03	455	AAQ75789 HS_2006.A2.E03.MR CIT	gb_ests8:W88219	7.00	116.25	1.1e+03	483	W88219 mf70c05.r1 Soares mou
gb_ests51:AAQ08630	7.00	116.66	1.0e+03	456	AAQ08630 SW05MCM37C11SK Onchd	gb_ests27:AA430352	7.00	116.25	1.1e+03	483	AA430352 mf71c10.y1 Sugano m
gb_ests52:AAQ75677	7.00	116.66	1.0e+03	456	AAQ75677 RPCI11-11P13.TP RPCI-11	gb_ests21:AAQ82967	7.00	116.23	1.1e+03	484	AAQ82967 uq24a04.r1 Soares_m
gb_ests53:AAQ61662	7.00	116.66	1.0e+03	456	AAQ61662 nbx0017C12r CUGI Rice	gb_ests23:AAQ08852	7.00	116.23	1.1e+03	484	AAQ08852 qb14a07.x1 Soares_p
gb_ests54:AAQ371205	7.00	116.64	1.0e+03	457	AAQ371205 HS_5048.A1.A08.SP6E RH	gb_ests27:AA409290	7.00	116.23	1.1e+03	484	AA409290 EST237582 Normalize
gb_ests55:AAQ1069	7.00	116.63	1.0e+03	458	AAQ1069 me31e08.r1 Soares mouse	gb_ests37:AA1979583	7.00	116.17	1.1e+03	488	AA1979583 sc89a06.y1 Gm-c1019
gb_ests56:AAQ53436	7.00	116.63	1.0e+03	458	AAQ53436 nf75g03.s1 NCI_CGAP_Cc	gb_ests23:AA116958	7.00	116.20	1.1e+03	486	AA116958 ue26a07.y1 Sugano m
gb_ests57:AAI211467	7.00	116.63	1.0e+03	458	AAI211467 p0d10a1.r1 Aspergillus	gb_ests36:AAI931202	7.00	116.20	1.1e+03	486	AAI931202 ul57f08.y1 Sugano m
gb_ests58:AAI308941	7.00	116.63	1.0e+03	458	AAI308941 q065d06.x1 NCI_CGAP_Cc	gb_ests16:AA531035	7.00	116.19	1.1e+03	487	AA531035 nk74b05.s1 NCI_CGAP
gb_ests59:AAI64639	7.00	116.63	1.0e+03	458	AAI64639 vq79a07.x1 Stratagene	gb_ests22:AAI037601	7.00	116.19	1.1e+03	487	AAI037601 oy31e01.x1 Soares_p
gb_ests60:AAQ56009	7.00	116.59	1.0e+03	460	AAQ56009 vk49e06.r1 Stratagene	gb_ests10:AA471007	7.00	116.17	1.1e+03	488	AA471007 ms54d08.r1 Life Tec
gb_ests61:AAI509975	7.00	116.58	1.0e+03	461	AAI509975 m115g03.y1 Soares mous	gb_ests4:AA437783	7.00	116.17	1.1e+03	488	AA437783 ve33e01.r1 Ko mouse
gb_ests62:AAI22851	7.00	116.58	1.0e+03	461	AAI22851 fc32e03.y1 Zebrafish W	gb_ests11:AAQ200128	7.00	116.16	1.1e+03	489	AAQ200128 mu10a05.r1 Soares_t
gb_ests63:AAI22937	7.00	116.58	1.0e+03	461	AAI22937 TENU4074.T.cruzi epim	gb_ests34:AAI800943	7.00	116.16	1.1e+03	489	AAI800943 wq06a05.x1 Soares_N
gb_ests64:AAQ71137	7.00	116.58	1.0e+03	461	AAQ71137 HS_5379.B2.G10.SP6E RH	gb_ests21:AA936235	7.00	116.16	1.1e+03	489	AA936235 on75b04.s1 Soares_N
gb_ests65:AAQ862698	7.00	116.56	1.0e+03	461	AAQ862698 nbx0019122f CUGI Rice	gb_ests45:AAW297946	7.00	116.13	1.1e+03	491	AAW297946 v25g07.r1 Soares_p
gb_ests66:AAQ17082	7.00	116.56	1.0e+03	462	AAQ17082 C72082 Rice panicle at d	gb_ests15:AA522038	7.00	116.12	1.1e+03	492	AA522038 vi08e1.r1 Barstead
gb_ests67:AAI59601	7.00	116.55	1.0e+03	463	AAI59601 z08a04.s1 Stratagene	gb_ests37:AAI979623	7.00	116.10	1.1e+03	493	AAI979623 614038C08.x2 614
gb_ests68:AAI63104	7.00	116.55	1.0e+03	463	AAI63104 mr93c11.r1 Stratagene	gb_gss16:AAQ621499	7.00	116.10	1.1e+03	493	AAQ621499 HS_3029.B2.F03.MF C
gb_ests69:AAQ060496	7.00	116.55	1.0e+03	463	AAQ060496 v137h10.x1 Stratagene	gb_ests27:AAI422091	7.00	116.09	1.1e+03	494	AAI422091 tf57g11.x1 NCI_CGAP
gb_ests70:AAI59751	7.00	116.53	1.0e+03	464	AAI59751 vq57a09.r1 Soares_mamm	gb_ests34:AAI802199	7.00	116.09	1.1e+03	494	AAI802199 tx25d07.x1 NCI_CGAP
gb_ests71:AAQ48858	7.00	116.53	1.0e+03	464	AAQ48858 UI-W-BHI-amf-b-12-O-UT	gb_ests37:AAI979583	7.00	116.09	1.1e+03	494	AAI979583 614036B02.x2 614
gb_ests72:AAQ129109	7.00	116.52	1.0e+03	465	AAQ129109 yH58c06.r1 Soares placen	gb_ests45:AAW289665	7.00	116.09	1.1e+03	494	AAW289665 NXNV004A04F.Nsf Xyl
gb_ests73:AAQ150271	7.00	116.49	1.0e+03	467	AAQ150271 vH58c04.r1 Soares_mamm	gb_gss15:AAQ551927	7.00	116.09	1.1e+03	494	AAQ551927 RPCI-11-42509.TJ RP
gb_ests74:AAQ1692133	7.00	116.49	1.0e+03	467	AAQ1692133 SW05MCM28E05SK Onchd	gb_gss16:AAQ65932	7.00	116.07	1.1e+03	495	AAQ65932 HS_5377.B2.A07.SP6E
gb_ests75:AAQ047580	7.00	116.49	1.0e+03	467	AAQ047580 UI-M-BHI-akm-b-01-O-UT	gb_gss14:AAQ72930	7.00	116.06	1.1e+03	496	AAQ72930 CITBI-11-2589A8.TR
gb_ests76:AAQ52797	7.00	116.49	1.0e+03	467	AAQ52797 614036E02.y1 614 - roc	gb_ests44:AAW249330	7.00	116.04	1.1e+03	497	AAW249330 AV393530 Chlamydomo
gb_ests77:AAQ259875	7.00	116.49	1.0e+03	467	AAQ259875 um79f05.y1 Sugano mous	gb_ests12:AAQ275949	7.00	116.03	1.1e+03	498	AAQ275949 CITBI-11-2524116.TR
gb_ests78:AAI25259	7.00	116.47	1.1e+03	468	AAI25259 mp71d03.r1 Soares_thym	gb_gss14:AAQ492751	7.00	116.03	1.1e+03	498	AAQ492751 HS_5219.B2.A07.SP6E

gb_est23:AI091239	7.00	116.01	1.1e+03	499	AI091239	oo26b01.x1 Soares_NSF	gb_est21:AA986571	7.00	115.46	1.2e+03	539	AA986571	uc81e06.y1 Sugano m
gb_est28:AI545064	7.00	116.01	1.1e+03	499	AI545064	fb70c05.y1 Zedrafish W	gb_est9:AA101694	7.00	115.45	1.2e+03	540	AA101694	rs01f12.r1 Sommer P
gb_est38:AI091742	7.00	116.01	1.1e+03	499	AI091742	ws5a12.x1 NCI_CGAP_BH	gb_est22:AI049371	7.00	115.45	1.2e+03	540	AI049371	uc83h12.y1 Sugano m
gb_est45:AW260594	7.00	116.01	1.1e+03	499	AW260594	um83f11.y1 NCI_CGAP	gb_est29:AI570810	7.00	115.45	1.2e+03	540	AI570810	tr67h12.x1 NCI_CGAP
gb_est10:AA412292	7.00	116.00	1.1e+03	500	AA412292	uc10g06.s1 Soares test	gb_gss4:AO697192	7.00	115.45	1.2e+03	540	AO697192	HS_5519.B2.A04.SP6E
gb_est20:AA898726	7.00	116.00	1.1e+03	500	AA898726	NCM6E10T7 Mycelial Neu	gb_est20:AA864874	7.00	115.44	1.2e+03	541	AA864874	oh03d08.s1 NCI_CGAP
gb_est15:AA524995	7.00	115.99	1.1e+03	501	AA524995	nh32c02.s1 NCI_CGAP_P	gb_est22:RI1894	7.00	115.40	1.2e+03	541	RI1894	yf50a09.r1 Soares Inf
gb_est23:AI098370	7.00	115.99	1.2e+03	501	AI098370	v984f11.r1 Barstead MH	gb_gss3:BB5816	7.00	115.39	1.2e+03	545	BB5816	CIT-HSP-2023C22.TF CI
gb_est44:AV394978	7.00	115.99	1.2e+03	501	AV394978	AV394978 Chlamydomonas	gb_est39:AA124037	7.00	115.37	1.2e+03	546	AA124037	EST292243 Normalized
gb_est14:AW222157	7.00	115.99	1.2e+03	501	AW222157	EST298968 tomato fruit	gb_gss12:AA128135	7.00	115.37	1.2e+03	546	AA128135	EST292243 Normalized
gb_gss16:AO671354	7.00	115.99	1.2e+03	501	AO671354	HS_5458.A2.C01.SP6E RH	gb_est21:C91180	7.00	115.36	1.2e+03	547	C91180	C91180 Dictyostelium
gb_est36:AI891634	7.00	115.94	1.2e+03	504	AI891634	ul57f08.x1 Sugano mouse	gb_gss1:PR0012887	7.00	115.36	1.2e+03	547	PR0012887	F rubripes GSS sequ
gb_gss8:AO015726	7.00	115.93	1.2e+03	505	AO015726	CIT-HSP-2307H1.TF CIT	gb_gss7:AO031965	7.00	115.36	1.2e+03	547	AO031965	RPC1-23-28F9.TV RP
gb_gss12:AAQ35831	7.00	115.93	1.2e+03	505	AAQ35831	HS_5016.B1.G12.SP6E RH	gb_gss5:AO827592	7.00	115.35	1.2e+03	548	AO827592	HS_5472.A2.F12.T7A
gb_gss12:AA672957	7.00	115.90	1.2e+03	507	AA672957	u48b03.r1 Barstead mous	gb_gss4:AO019132	7.00	115.33	1.3e+03	549	AO019132	HS_5396.B2.E02.T7A
gb_est25:AA1317173	7.00	115.90	1.2e+03	507	AA1317173	u133h09.y1 Soares mous	gb_gss13:AAQ430015	7.00	115.32	1.3e+03	550	AAQ430015	HS_5061.B2.D02.T7A
gb_est8:AA009462	7.00	115.89	1.2e+03	508	AA009462	ze82q01.r1 Soares feta	gb_est17:AA617054	7.00	115.31	1.3e+03	551	AA617054	v121c04.r1 Barstead
gb_est8:AA009462	7.00	115.89	1.2e+03	508	AA009462	ze82q01.r1 Soares feta	gb_est23:AI088606	7.00	115.31	1.3e+03	551	AI088606	vb14e01.x1 Soares_P
gb_est29:AI607620	7.00	115.89	1.2e+03	508	AI607620	vk49e06.y1 Stratagene	gb_est37:AI976931	7.00	115.31	1.3e+03	551	AI976931	EST271468 Schistos
gb_est36:AI901418	7.00	115.89	1.2e+03	508	AI901418	ul61c08.y1 Sugano mouse	gb_est10:AA167479	7.00	115.29	1.3e+03	552	AA167479	zp09h11.r1 Stratage
gb_est34:AI9017071	7.00	115.89	1.2e+03	508	AI9017071	ul92f02.y1 Sugano mouse	gb_gss11:AA189784	7.00	115.29	1.3e+03	552	AA189784	HS_3224.A2.E12.T7 C
gb_est29:AI607665	7.00	115.87	1.2e+03	509	AI607665	va86h11.y1 Soares mouse	gb_gss14:AAQ493795	7.00	115.29	1.3e+03	552	AAQ493795	HS_5071.A1.H12.T7 R
gb_est37:AI994270	7.00	115.87	1.2e+03	509	AI994270	701501945.A. thaliana	gb_est25:AI279395	7.00	115.28	1.3e+03	553	AI279395	qx66b07.x1 NCI_CGAP
gb_gss6:AO858570	7.00	115.87	1.2e+03	509	AO858570	nbe0002B1r CUGI Rice	gb_gss13:AAQ357745	7.00	115.27	1.3e+03	554	AAQ357745	CITB1-EL-2537H13.TF
gb_est11:AA271129	7.00	115.86	1.2e+03	510	AA271129	va86h11.r1 Soares mouse	gb_est32:AI734529	7.00	115.26	1.3e+03	555	AI734529	606032F08.x1 606
gb_gss1:CMS001IE	7.00	115.86	1.2e+03	510	CMS001IE	Arabidopsis thaliana g	gb_gss3:BB68832	7.00	115.26	1.3e+03	555	BB68832	CIT-HSP-2024B19.TR CI
gb_gss4:AO687728	7.00	115.86	1.2e+03	510	AO687728	nbs005F09r CUGI Rice	gb_est22:AI006180	7.00	115.23	1.3e+03	557	AI006180	uaf7g03.r1 Soares_m
gb_gss10:AAQ132220	7.00	115.86	1.2e+03	510	AAQ132220	HS_3005.A1.E10.MF CIT	gb_est37:AI939456	7.00	115.23	1.3e+03	557	AI939456	tf19f01.x1 NCI_CGAP
gb_est21:AA939273	7.00	115.85	1.2e+03	511	AA939273	oq31b07.s1 NCI_CGAP GC	gb_est38:AA049416	7.00	115.22	1.3e+03	558	AA049416	UI-N-BH1-ane-q-12-0
gb_gss3:BB27068	7.00	115.85	1.2e+03	511	BB27068	T27F7F TAMU Arabidopsis	gb_est27:AI414985	7.00	115.20	1.3e+03	559	AI414985	vb46c10.x1 Soares m
gb_gss13:AAQ39533	7.00	115.85	1.2e+03	511	AAQ39533	HS_5033.B2.B08.T7 RPT	gb_est29:AI586070	7.00	115.20	1.3e+03	559	AI586070	vu91c11.x1 Stratage
gb_gss16:AO620640	7.00	115.85	1.2e+03	511	AO620640	HS_5183.B2.C04.SP6E RH	gb_est31:AA676908	7.00	115.20	1.3e+03	559	AA676908	we57f06.x1 Soares_t
gb_est35:AI854072	7.00	115.83	1.2e+03	512	AI854072	UI-N-BH0-ajm-h-01-0-UI	gb_est21:AA986204	7.00	115.18	1.3e+03	561	AA986204	uc80a03.y1 Sugano m
gb_gss15:AI869290	7.00	115.80	1.2e+03	514	AI869290	ul17g01.y1 Sugano mouse	gb_est32:AI745947	7.00	115.17	1.3e+03	562	AI745947	605077G12.x1 605
gb_est38:AO892289	7.00	115.80	1.2e+03	514	AO892289	EST285469 tomato mixed	gb_est44:AAW173310	7.00	115.17	1.3e+03	562	AAW173310	xj85h03.x1 Soares_N
gb_est44:AAW206294	7.00	115.80	1.2e+03	514	AAW206294	UI-N-BH1-afe-q-08-0-UI	gb_gss7:AO950822	7.00	115.17	1.3e+03	562	AO950822	Sheared DNA-5C20.T
gb_gss15:AAQ35660	7.00	115.80	1.2e+03	514	AAQ35660	RPC1-11-435f11.TV RPT	gb_est15:AA470014	7.00	115.15	1.3e+03	563	AA470014	u49f004.r1 Soares_f
gb_est30:AI683457	7.00	115.79	1.2e+03	515	AI683457	uc33a11.y1 Sugano mouse	gb_est23:AI097495	7.00	115.15	1.3e+03	563	AI097495	qb90f04.x1 Soares_f
gb_est32:AI099734	7.00	115.76	1.2e+03	517	AI099734	TENU2778 T. cruzi epln	gb_gss13:AAQ74683	7.00	115.14	1.3e+03	564	AAQ74683	EST255783 tomato re
gb_est36:AI929639	7.00	115.76	1.2e+03	517	AI929639	au61h05.y1 Schneider f	gb_gss7:AO925711	7.00	115.14	1.3e+03	564	AO925711	RPC1-23-297011.TJ R
gb_est36:AI912708	7.00	115.73	1.2e+03	518	AI912708	wel3g12.x1 NCI_CGAP_L	gb_gss11:AAQ156765	7.00	115.12	1.3e+03	566	AAQ156765	nbs0008K18f CUGI R
gb_est30:AA861815	7.00	115.73	1.2e+03	519	AA861815	ak35g07.x1 Soares test	gb_gss15:AAQ51466	7.00	115.12	1.3e+03	566	AAQ51466	HS_5413.A1.H02.T7A
gb_est16:AAW370475	7.00	115.73	1.2e+03	519	AAW370475	RC2-ET0256-111199-012	gb_gss16:AA341729	7.00	115.09	1.3e+03	568	AA341729	ni99d02.s1 NCI_CGAP
gb_est33:AI772785	7.00	115.72	1.2e+03	520	AI772785	EST233885 tomato resis	gb_est9:AA069964	7.00	115.08	1.3e+03	569	AA069964	mm61f10.r1 Stratage
gb_est36:AI931250	7.00	115.72	1.2e+03	520	AI931250	ul58g08.y1 Sugano mouse	gb_est17:C77166	7.00	115.08	1.3e+03	569	C77166	C77166 Mouse 3.5-dpc
gb_est36:AI931644	7.00	115.72	1.2e+03	520	AI931644	ul17g01.y1 Sugano mouse	gb_est18:AA726464	7.00	115.05	1.3e+03	571	AA726464	v460e10.r1 Barstead
gb_est36:AI931208	7.00	115.71	1.2e+03	521	AI931208	ul57g03.y1 Sugano mouse	gb_est37:AI977956	7.00	115.05	1.3e+03	571	AI977956	496036G01.x1 496
gb_gss4:AO703085	7.00	115.71	1.2e+03	521	AO703085	HS_5466.A2.F10.T7A RPT	gb_est30:AA048988	7.00	115.03	1.3e+03	573	AA048988	DKF2P43A1N1018.s1 43
gb_gss5:AO783295	7.00	115.69	1.2e+03	521	AO783295	HS_5445.A2.E10.T7A RPT	gb_gss8:AO978919	7.00	115.03	1.3e+03	573	AO978919	RPC1-23-339A6.TV RP
gb_est34:AI806813	7.00	115.69	1.2e+03	522	AI806813	wf15f03.x1 Soares_NFL	gb_gss13:AAQ398831	7.00	115.02	1.3e+03	574	AAQ398831	mgx00015C01f CUGI R
gb_est46:AAW384748	7.00	115.69	1.2e+03	522	AAW384748	un05g05.y1 Sugano mouse	gb_est39:AAW106921	7.00	114.99	1.3e+03	576	AAW106921	um34c05.x1 Sugano m
gb_gss11:AAQ193115	7.00	115.69	1.2e+03	522	AAQ193115	HS_3199.B1.A11.T7 CIT	gb_est25:AI039768	7.00	114.98	1.3e+03	577	AI039768	qo75b03.x1 NCI_CGAP
gb_gss15:AAQ33680	7.00	115.67	1.2e+03	523	AAQ33680	RPC1-11-384K11.TV RPT	gb_est32:AAQ76669	7.00	114.98	1.3e+03	577	AAQ76669	BNLGH15789.S1x-day
gb_est18:R211750	7.00	115.67	1.2e+03	524	R21750	3758 Lambda-PRL2 Arabid	gb_gss13:AAQ368594	7.00	114.98	1.3e+03	577	AAQ368594	HS_5037.A2.G07.SP6E
gb_gss11:AAQ175573	7.00	115.67	1.2e+03	524	AAQ175573	HS_3215.A2.F12.T7 CIT	gb_gss13:AAQ300093	7.00	114.98	1.3e+03	577	AAQ300093	RPC111-146E23.TV RP
gb_est21:AAQ86570	7.00	115.65	1.2e+03	525	AAQ86570	uc81e05.y1 Sugano mouse	gb_est38:AAW086627	7.00	114.97	1.3e+03	578	AAW086627	2821e12.x1 Normal NH
gb_gss11:AAQ216796	7.00	115.65	1.2e+03	525	AAQ216796	HS_2262.A1.D01.MR CIT	gb_est34:AAW106921	7.00	114.95	1.3e+03	578	AAW106921	2821e19.Sprime NIH
gb_est38:AI540285	7.00	115.64	1.2e+03	526	AI540285	tg34a09.x1 NCI_CGAP_P	gb_gss12:AAQ259708	7.00	114.95	1.3e+03	579	AAQ259708	nbs0032L09r CUGI R
gb_gss10:AAQ164481	7.00	115.63	1.2e+03	527	AAQ164481	HS_3007.B2.C01.T7 CIT	gb_est15:AA018350	7.00	114.94	1.3e+03	580	AA018350	v110b12.y1 Barstead
gb_gss11:AAQ23487	7.00	115.61	1.2e+03	528	AAQ23487	HS_2003.B2.G01.MR CIT	gb_est17:AA628539	7.00	114.92	1.3e+03	582	AA628539	af27h06.s1 Soares_t
gb_est15:AA517310	7.00	115.60	1.2e+03	529	AA517310	vh86h04.r1 Knowles Sol	gb_est37:AI963374	7.00	114.92	1.3e+03	582	AI963374	ul74d12.y1 Sugano m
gb_est28:AI546582	7.00	115.60	1.2e+03	529	AI546582	u440c09.y1 Sugano mouse	gb_gss12:AAQ73510	7.00	114.92	1.3e+03	582	AAQ73510	nbs0030007r CUGI R
gb_gss16:AO603724	7.00	115.60	1.2e+03	529	AO603724	HS_2116.A2.G09.T7C CIT	gb_est24:AAQ166936	7.00	114.91	1.3e+03	583	AAQ166936	xylem.est-17134021.PJ R
gb_est37:AI400162	7.00	115.58	1.2e+03	530	AI400162	tg67g09.y1 Soares NHM	gb_gss13:AAQ368610	7.00	114.91	1.3e+03	583	AAQ368610	rpcc11-134021.TJ R
gb_est30:AA869046	7.00	115.57	1.2e+03	531	AA869046	Q30b01.r1 Barstead st	gb_est7:WF65917	7.00	114.89	1.3e+03	584	WF65917	me09d09.r1 Soares mou
gb_est32:AAQ394579	7.00	115.54	1.2e+03	533	AAQ394579	CITB1-ER-2353H20.TR CIT	gb_est24:AA128923	7.00	114.88	1.3e+03	585	AA128923	EST225618 Normalized
gb_est11:AA286644	7.00	115.53	1.2e+03	534	AA286644	uk21d012.r1 Soares mouse	gb_gss7:AO901624	7.00	114.88	1.3e+03	585	AO901624	HS_2014.A2.H01.MR C
gb_est30:AI648233	7.00	115.53	1.2e+03	534	AI648233	mr4d012.x1 Sugano mouse	gb_gss15:AAQ54295	7.00	114.87	1.3e+03	586	AAQ54295	RPC1-11-422M1.TV R
gb_est39:AAW198873	7.00	115.53	1.2e+03	534	AAW198873	xe89c02.x1 NCI_CGAP_B	gb_est19:AA125356	7.00	114.86	1.3e+03	587	AA125356	vp98a12.r1 Soares_m

gb_est38:AW05233	7.00	114.78	1.3e+03	593	AW05233	wz02f10.x1 NCI_CGAP_B4	gb_gss12:AO287753	7.00	113.85	1.5e+03	676	AQ287753	nbxb0031C14r CUGI R
gb_est10:AA173580	7.00	114.75	1.3e+03	596	AA173580	z04f03.s1 Stratagene	gb_gss1:AG004174	7.00	113.82	1.5e+03	679	AG004174	Homo sapiens genomi
gb_est45:AW09235	7.00	114.75	1.3e+03	596	AW09235	sf29b03.x1 Gm-c1028 Cl	gb_gss1:AG010300	7.00	113.82	1.5e+03	679	AG010300	Homo sapiens genomi
gb_est47:AW042692	7.00	114.75	1.3e+03	596	AW042692	EST307622 tm-comto mixed	gb_est46:AW318858	7.00	113.79	1.5e+03	682	AW318858	un08e08.y1 Sugano m
gb_gss3:AB28252	7.00	114.75	1.3e+03	596	B28252	T7G13PFB TAMU Arabidopsi	gb_est46:AW341276	7.00	113.77	1.5e+03	684	AW341276	x297d10.x1 NCI_CGAP
gb_gss13:AO403772	7.00	114.75	1.3e+03	596	AO403772	HS_5049.A1.A06.SP6E RF	gb_est44:AW200005	7.00	113.76	1.5e+03	685	AW200005	da10d10.x1 Xenopus
gb_gss12:AO329142	7.00	114.74	1.4e+03	597	AO329142	nbxb0044024f CUGI Rice	gb_gss12:AO487682	7.00	113.73	1.5e+03	687	AQ487682	RPCI-11-362C3.TJ Rp
gb_est39:AW106842	7.00	114.72	1.4e+03	598	AW106842	un34c05.y1 Sugano mous	gb_est26:AW0025919	7.00	113.71	1.5e+03	689	AW0025919	AB002449 Bombyx mor
gb_gss16:AO606757	7.00	114.72	1.4e+03	598	AO606757	HS_5401.B1.F01.T7A RPC	gb_gss11:AO255949	7.00	113.70	1.5e+03	690	AQ255949	nbxb0014F21r CUGI R
gb_est10:AA130250	7.00	114.70	1.4e+03	600	AA130250	a138h11.s1 Soares pred	gb_gss1:AG0014351	7.00	113.67	1.5e+03	693	AG0014351	Homo sapiens genomi
gb_est16:AA543360	7.00	114.69	1.4e+03	601	AA543360	rj81b12.r1 Soares mam	gb_gss1:AG001298	7.00	113.67	1.5e+03	693	AG001298	Homo sapiens genomi
gb_est12:AA276254	7.00	114.69	1.4e+03	601	AA276254	vc40f06.r1 Barstead MR	gb_est21:AA942252	7.00	113.62	1.6e+03	698	AA942252	MD26355.5Prime LD D
gb_gss2:FR0036512	7.00	114.68	1.4e+03	602	FR0036512	AL124021 Fugu rubripes GSS sequ	gb_gss10:AO656770	7.00	113.61	1.6e+03	699	AQ656770	Sheared DNA-21D17.T
gb_gss3:BB3768	7.00	114.68	1.4e+03	602	BB3768	CPG0107B CpiOwAGNAL Crx	gb_gss16:AO656770	7.00	113.61	1.6e+03	699	AQ656770	Sheared DNA-21D17.T
gb_est35:AI958999	7.00	114.65	1.4e+03	604	AI958999	w132b12.x1 NCI_CGAP_U	gb_gss12:AO634693	7.00	113.60	1.6e+03	700	AQ634693	mgxb0019D15r CUGI R
gb_gss15:AO561834	7.00	114.65	1.4e+03	604	AO561834	HS_2094.A2.H09 MR CIT	gb_est38:AW054379	7.00	113.55	1.6e+03	705	AW054379	L30-17173 Ice plan
gb_gss4:AO73418	7.00	114.64	1.4e+03	605	AO73418	HS_5486.A1.F02.T7A RPC	gb_est34:AI790807	7.00	113.53	1.6e+03	707	AI790807	wt28c03.y1 Sugano m
gb_gss11:AA113913	7.00	114.63	1.4e+03	606	AA113913	zm80h09.r1 Stratagene	gb_est47:AW0424873	7.00	113.53	1.6e+03	707	AW0424873	ESTPM004.Penaeus m
gb_est9:AA110381	7.00	114.61	1.4e+03	608	AA110381	mo26f11.r1 Life Tech m	gb_gss3:BB77441	7.00	113.52	1.6e+03	708	BB77441	T31A8TFC TAMU Arabido
gb_gss7:AO926169	7.00	114.61	1.4e+03	608	AO926169	RPCI-23-27909.TV RPCI-1	gb_est47:AW422328	7.00	113.51	1.6e+03	709	AW422328	f161a10.y1 Sugano K
gb_gss7:AO926169	7.00	114.59	1.4e+03	609	AO926169	RPCI-23-27909.TV RPCI-1	gb_gss13:AO382587	7.00	113.51	1.6e+03	709	AQ382587	RPCI11-139C16.TJ RP
gb_est23:AI114181	7.00	114.58	1.4e+03	610	AI114181	GH10889.5Prime GH Dros	gb_est24:AO032717	7.00	113.49	1.6e+03	711	AO032717	AO032717 Rice green
gb_est47:AW10145	7.00	114.58	1.4e+03	610	AW10145	fh04e01.x1 NIH MGC.17	gb_gss1:AG001326	7.00	113.44	1.6e+03	716	AG001326	Homo sapiens genomi
gb_gss12:AO271066	7.00	114.58	1.4e+03	610	AO271066	nbxb0008B09f CUGI Rice	gb_gss9:AO051495	7.00	113.43	1.6e+03	717	AO051495	nbxb0002b011r CUGI
gb_gss12:AO271066	7.00	114.56	1.4e+03	612	AO271066	L30-248573 Ice plant I	gb_gss14:AO478732	7.00	113.42	1.6e+03	718	AQ478732	RPCI-11-268L14.TJ R
gb_est46:AW265953	7.00	114.55	1.4e+03	613	AW265953	707016D01.x1 707 - Mix	gb_gss1:AG001525	7.00	113.39	1.6e+03	721	AG001525	Homo sapiens genomi
gb_gss1:FR000992	7.00	114.55	1.4e+03	613	FR000992	AL001174 F. rubripes GSS sequ	gb_gss6:AO865465	7.00	113.39	1.6e+03	721	AO865465	nbbe0025K14f CUGI R
gb_gss16:AO620477	7.00	114.55	1.4e+03	613	AO620477	HS_5186.B1.B04.T7A RPC	gb_gss7:AO957386	7.00	113.38	1.6e+03	722	AO957386	LERAP48TFL LERA Arab
gb_est14:AA427126	7.00	114.52	1.4e+03	615	AA427126	vc24b03.r1 Soares mous	gb_gss1:AG014310	7.00	113.37	1.6e+03	723	AG014310	Homo sapiens genomi
gb_gss12:AO271066	7.00	114.52	1.4e+03	615	AO271066	nbxb001711r CUGI Rice	gb_gss1:AG014310	7.00	113.37	1.6e+03	723	AG014310	Homo sapiens genomi
gb_gss12:AO271066	7.00	114.51	1.4e+03	615	AO271066	RPCI11-1F5.TJ RPCI-11	gb_gss1:AG014310	7.00	113.35	1.6e+03	725	AG014310	Homo sapiens genomi
gb_gss12:AO271066	7.00	114.51	1.4e+03	615	AO271066	RPCI11-1F5.TJ RPCI-11	gb_gss1:AG014310	7.00	113.35	1.6e+03	725	AG014310	Homo sapiens genomi
gb_est16:AA105033	7.00	114.51	1.4e+03	616	AA105033	uc383d11.y1 Sugano mous	gb_gss14:AO474351	7.00	113.34	1.6e+03	726	AO474351	CITBI-EL-2595A16.TR
gb_est36:AI876754	7.00	114.51	1.4e+03	616	AI876754	u33e04.y1 Sugano mous	gb_gss14:AO474351	7.00	113.34	1.6e+03	726	AO474351	CITBI-EL-2595A16.TR
gb_est19:AA082096	7.00	114.49	1.4e+03	618	AA082096	GM06330.5Prime GM Dros	gb_est28:AI527764	7.00	113.27	1.6e+03	733	AI527764	u28g05.y1 Sugano m
gb_gss1:FR000985	7.00	114.48	1.4e+03	619	FR000985	EST246391 tomato ovary	gb_gss4:AO729408	7.00	113.27	1.6e+03	733	AO729408	HS_5570.B1.E12.SP6
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss9:AO054807	7.00	113.27	1.6e+03	733	AO054807	CLM-7ell-u CLM Glar
gb_gss9:AO08301	7.00	114.48	1.4e+03	619	AO08301	CPG0433A CpiOwAGNAL C	gb_gss						

gb_est23:AI114963	7.00	112.40	1.8e+03	828	AI114963 u14all.y1 Sugano mous	gb_gss12:AO341247	6.00	107.39	3.5e+03	111	AQ341247 RPCI11-11N21.TV RP
gb_gss3:BI9885	7.00	112.36	1.8e+03	833	BI9885 T18N7-Sp6 TAMU Arabidops	gb_est8:AA014128	6.00	107.32	3.5e+03	112	AA014128 mh28h01.r1 Soares m
gb_est34:AI789173	7.00	112.35	1.8e+03	834	AI789173 uK51h07.y1 Sugano mous	gb_est15:AA484507	6.00	107.32	3.5e+03	112	AA484507 nf08f01.s1 NCI_CGAP
gb_gss5:AO780692	7.00	112.33	1.8e+03	837	AO780692 HS_3169_A1_C04-77C CIT	gb_est18:AA711354	6.00	107.32	3.5e+03	112	AA711354 vt77b08.r1 Barstead
gb_gss1:CN5009EQ	7.00	112.25	1.9e+03	846	AO53482 Drosophila melanogaste	gb_est26:AI3166474	6.00	107.32	3.5e+03	112	AI3166474 ao83d11.x1 Stanley
gb_gss5:AO493772	7.00	112.24	1.9e+03	847	AO493772 HS_5573_A2_E05-w7A RPD	gb_est45:AW259612	6.00	107.32	3.5e+03	112	AW259612 up31a12.x1 NCI_CGAP
gb_gss6:AO853523	7.00	112.19	1.9e+03	853	AO853523 nbx2002IC20f CUGI Rice	gb_est15:AA484423	6.00	107.26	3.5e+03	113	AA484423 nf07b04.s1 NCI_CGAP
gb_est28:AI327753	7.00	112.14	1.9e+03	859	AI327753 y1 Sugano mous	gb_est15:AA4844509	6.00	107.26	3.5e+03	113	AA4844509 nf08f04.s1 NCI_CGAP
gb_gss15:AO576187	7.00	112.03	1.9e+03	873	AO576187 nbx00808P23f CUGI Rice	gb_est6:NA6349	6.00	107.20	3.6e+03	114	NA6349 J7848F Human fetal he
gb_gss38:AW054104	7.00	112.00	1.9e+03	876	AW054104 L30-2085T3 Ice plant I	gb_gss9:AO042290	6.00	107.13	3.6e+03	115	AO042290 CIT-HSP-2330G3.TF C
gb_est31:AO067178	7.00	111.96	1.9e+03	881	AO067178 AO67178 Sugano mouse	gb_est5:DI1817	6.00	107.07	3.6e+03	116	DI1817 MUGSG00362 Mouse live
gb_gss1:CN5005IB	7.00	111.90	1.9e+03	882	AO059336 Drosophila melanogaste	gb_est19:AA575093	6.00	107.07	3.6e+03	116	AA575093 ah55e10.s1 Soares_t
gb_est25:AI323216	7.00	111.71	2.0e+03	913	AI323216 mp1e08.y1 Soares_thym	gb_est22:AI015107	6.00	107.07	3.6e+03	116	AI015107 ct90b12.s1 Soares_t
gb_gss4:AO688009	7.00	111.66	2.0e+03	919	AO688009 nbx0070F08r CUGI Rice	gb_gss3:R22231	6.00	107.01	3.6e+03	117	R22231 Yt93g10.s1 Soares_pln
gb_gss3:BI1110	7.00	111.65	2.0e+03	920	BI1110 T7G13-T7 TAMU Arabidops1	gb_est15:AA584435	6.00	107.01	3.6e+03	117	B23317 F18J6TF IGF Arabidops
gb_gss1:CN5009BQC	7.00	111.39	2.1e+03	954	AO57502 Drosophila melanogaste	gb_est16:AA549126	6.00	106.95	3.7e+03	118	AA549126 nf07b06.s1 NCI_CGAP
gb_gss6:AO895470	7.00	111.30	2.1e+03	966	AO895470 HS_3117_A2_A01_MR CIT	gb_est18:AA736673	6.00	106.95	3.7e+03	118	AA736673 oa22e06.s1 NCI_CGAP
gb_gss2:CN50188V	7.00	111.21	2.1e+03	979	AI109033 Drosophila melanogaste	gb_est5:U45337	6.00	106.89	3.7e+03	119	U45337 HUMHG15111 Human cere
gb_gss2:NA0557	7.00	111.14	2.1e+03	989	NA0557 mc34ell.r1 Soares mouse	gb_est15:AA501682	6.00	106.89	3.7e+03	119	AA501682 ng06e03.s1 NCI_CGAP
gb_gss2:CN5017C9	7.00	110.87	2.2e+03	1027	AI107859 Drosophila melanogaste	gb_est20:AA866045	6.00	106.89	3.7e+03	119	AA866045 oh52g08.s1 NCI_CGAP
gb_gss1:CN5007MA	7.00	110.72	2.3e+03	1049	AO67244 Drosophila melanogaste	gb_gss15:AO585855	6.00	106.89	3.7e+03	119	AO585855 RPCI-11-432E1.TJ RP
gb_gss1:CN5000FF	7.00	110.37	2.4e+03	1101	AL052862 Drosophila melanogaste	gb_est16:AA596402	6.00	106.83	3.7e+03	120	AA596402 vo38d09.r1 Barstead
gb_gss1:CN5000NB	7.00	110.37	2.4e+03	1101	AL063259 Drosophila melanogaste	gb_est46:AW347413	6.00	106.83	3.7e+03	120	AW347413 31109 MARC 2P1G Sus
gb_gss1:CN5001IX	7.00	110.37	2.4e+03	1101	AL061064 Drosophila melanogaste	gb_gss7:AO34158	6.00	106.83	3.7e+03	120	AO34158 RPCI-23-331H10.TJ R
gb_gss1:CN50024A	7.00	110.37	2.4e+03	1101	AL061833 Drosophila melanogaste	gb_gss9:AO036587	6.00	106.83	3.7e+03	120	AO036587 CIT-HSP-234108.TF C
gb_gss1:CN5002R2	7.00	110.37	2.4e+03	1101	AL063292 Drosophila melanogaste	gb_gss9:AO077099	6.00	106.83	3.7e+03	120	AO077099 CIT-HSP-2359F11.Tr
gb_gss1:CN50033P	7.00	110.37	2.4e+03	1101	AL063815 Drosophila melanogaste	gb_est3:R53268	6.00	106.77	3.8e+03	121	R53268 Y982A03.r1 Soares inf
gb_gss1:CN5003E3V	7.00	110.37	2.4e+03	1101	AL068708 Drosophila melanogaste	gb_est17:AA642376	6.00	106.77	3.8e+03	121	AA642376 ns25e03.s1 NCI_CGAP
gb_gss1:CN5000K4Z	7.00	110.37	2.4e+03	1101	AL077231 Drosophila melanogaste	gb_gss7:AO930968	6.00	106.77	3.8e+03	121	AO930968 RPCI-23-255F5.TJ RP
gb_gss3:BI0533	7.00	109.98	2.5e+03	1163	BI0533 F1J32-Sp6 IGF Arabidops1	gb_est15:AA484584	6.00	106.71	3.8e+03	122	AA484584 nf10b02.s1 NCI_CGAP
gb_gss3:BI1549	6.00	109.71	2.6e+03	1208	BI1549 F2689-Sp6 IGF Arabidops	gb_est15:AA503773	6.00	106.71	3.8e+03	122	AA503773 ng07h06.s1 NCI_CGAP
gb_est24:AI5221747	6.00	115.82	1.2e+03	34	AI5221747 q993f02.x1 Soares_NFL.T	gb_gss3:R35983	6.00	106.71	3.8e+03	122	R35983 HS-1031-A2-B11-MR.ab1
gb_est9:AA124006	6.00	113.22	1.6e+03	49	AA124006 mq10a06.r1 Barstead MPI	gb_est37:AA950826	6.00	106.66	3.8e+03	123	AA950826 wx55f09.x1 NCI_CGAP
gb_est19:AA57476	6.00	113.22	1.6e+03	49	AA57476 z941c04.s1 Soares_pinea	gb_gss3:R82206	6.00	106.66	3.8e+03	123	R82206 RPCI11-13C4.TVB RPCI-
gb_est46:AW318967	6.00	113.07	1.7e+03	50	AW318967 n10b04.y1 Sugano mouse	gb_est17:AA611735	6.00	106.60	3.8e+03	124	AA611735 vo89g03.r1 Barstead
gb_est45:AW255471	6.00	112.14	1.9e+03	57	AW255471 xq63f03.x1 NCI_CGAP_NH7	gb_est17:AA020681	6.00	106.60	3.8e+03	124	AA020681 af89b06.s1 Soares_t
gb_est18:AA930665	6.00	112.02	1.9e+03	58	AA930665 n449g03.x1 NCI_CGAP_EW1	gb_est20:AA852135	6.00	106.60	3.8e+03	124	AA852135 NHTCag10602f1 Norm
gb_est11:AA240825	6.00	111.43	2.1e+03	63	AA240825 mw17e06.r1 Soares mouse	gb_est20:AA902765	6.00	106.60	3.8e+03	124	AA902765 ok72a01.s1 NCI_CGAP
gb_est21:AA240825	6.00	111.43	2.1e+03	63	AA240825 mw17e06.r1 Soares mouse	gb_est25:AI321028	6.00	106.60	3.8e+03	124	AI321028 d2h07nm.r1 Neurospo
gb_est20:AA822882	6.00	110.99	2.2e+03	67	AA822882 oJ89c03.s1 Soares_NFL.T	gb_est8:AA306515	6.00	106.54	3.9e+03	125	AA306515 m159g11.r1 Soares m
gb_est20:AA822882	6.00	110.67	2.3e+03	70	AA822882 qj91g12.r1 Knowles Solt	gb_est16:AA545426	6.00	106.48	3.9e+03	126	AA545426 vj32d06.x1 NCI_CGAP
gb_est18:AA716326	6.00	110.18	2.4e+03	75	AA716326 z961g02.s1 Soares_fetal	gb_est13:AI19204	6.00	106.48	3.9e+03	126	AI19204 t32d06.x1 NCI_CGAP
gb_est14:AA13641	6.00	110.00	2.5e+03	77	AA13641 vc57b02.s1 Knowles Solt	gb_est44:AA392280	6.00	106.48	3.9e+03	126	AA392280 AV392280 Chlanydomo
gb_est4:AA40990	6.00	109.90	2.5e+03	78	AA40990 yn96d02.s1 Soares adult b	gb_gss13:AA372710	6.00	106.48	3.9e+03	126	AA372710 RPCI11-157E2.TV RPC
gb_est8:AA030267	6.00	109.81	2.5e+03	79	AA030267 m126b06.r1 Soares mouse	gb_est10:AA152650	6.00	106.43	3.9e+03	127	AA152650 mq55c07.r1 Soares t
gb_est30:AI663950	6.00	109.81	2.5e+03	79	AI663950 uJ46h04.x1 Sugano mouse	gb_est15:AA484382	6.00	106.43	3.9e+03	127	AA484382 nf06f11.s1 NCI_CGAP
gb_est44:AW191390	6.00	109.72	2.6e+03	80	AW191390 suc204 Eucalypt cambium	gb_gss9:AO073803	6.00	106.43	3.9e+03	127	AO073803 EP(3)3186-3prime Dr
gb_est25:AI510837	6.00	109.46	2.7e+03	83	AI510837 ta43a11.x2 NCI_CGAP_Ju2	gb_est6:NA61751	6.00	106.37	4.0e+03	128	NA61751 T952zy29f02.r1 TgrH
gb_est4:AA53369	6.00	109.21	2.7e+03	86	AA53369 CHR220308 Chromosome 22 e	gb_est10:AA167932	6.00	106.37	4.0e+03	128	AA167932 ms21d08.r1 Stratage
gb_est14:AA433749	6.00	108.80	2.9e+03	91	AA433749 vf56b01.r1 Soares mouse	gb_est15:AA484594	6.00	106.37	4.0e+03	128	AA484594 nf10c06.s1 NCI_CGAP
gb_est25:AI313603	6.00	108.80	2.9e+03	91	AI313603 EST057 Mouse ES cell La	gb_est16:AA557247	6.00	106.37	4.0e+03	128	AA557247 nl75e08.s1 NCI_CGAP
gb_est15:AA529216	6.00	108.65	3.0e+03	93	AA529216 v133f10.r1 Barstead mou	gb_est21:AA973289	6.00	106.37	4.0e+03	128	AA973289 on94b04.s1 Soares_N
gb_est27:AI407374	6.00	108.65	3.0e+03	93	AI407374 EST935663 Normalized ra	gb_est31:AI708130	6.00	106.37	4.0e+03	128	AI708130 as99h10.x1 Barstead
gb_est11:AA207832	6.00	108.57	3.0e+03	94	AA207832 mv73h04.r1 Guaywoodford	gb_est28:AI506278	6.00	106.32	4.0e+03	129	AI506278 vq55f01.x1 Barstead
gb_est21:AA910331	6.00	108.42	3.0e+03	96	AA910331 ok83f04.s1 NCI_CGAP_Kid	gb_est11:AA392280	6.00	106.26	4.0e+03	130	AA392280 hfe0419.seq.F Human
gb_est29:AA092430	6.00	108.35	3.1e+03	97	AA16225 f2b0b04.y1 Zebrafish wa	gb_est16:NA87868	6.00	106.21	4.0e+03	131	NA87868 K040F Human fetal he
gb_est40:AA092430	6.00	108.13	3.2e+03	100	AA092430 119427.seq.F Human fet	gb_est6:NA62011	6.00	106.15	4.1e+03	132	NA62011 T952zy29f06.r1 TgrH
gb_est41:AA092430	6.00	108.13	3.2e+03	100	AA092430 119427.seq.F Human fet	gb_est15:AA469337	6.00	106.15	4.1e+03	132	AA469337 nc70y207.s1 NCI_CGAP
gb_est37:AI959589	6.00	108.06	3.2e+03	101	AI959589 z62a04.s3 Xenopus EST	gb_est30:AI505039	6.00	106.15	4.1e+03	132	AI505039 wa18f10.x1 NCI_CGAP
gb_gss9:AO603074	6.00	107.92	3.2e+03	103	AO603074 wt89a07.x1 NCI_CGAP_Me	gb_est31:AV029631	6.00	106.15	4.1e+03	132	AV029631 AV029631 Mus muscul
gb_est4:AA55145	6.00	107.85	3.3e+03	104	AA55145 CHR220084 Chromosome 22	gb_est20:AA834554	6.00	106.10	4.1e+03	133	AA834554 z483g04.s1 Soares_p
gb_gss9:AO60926	6.00	107.85	3.3e+03	104	AO60926 CIT-HSP-2356h10.TF CIT	gb_est23:AO97861	6.00	106.10	4.1e+03	133	AO97861 ue40g07.y1 Sugano m
gb_est6:AO6645	6.00	107.78	3.3e+03	105	AO6645 T952zy29f06.s1 Stanley	gb_est40:AW151370	6.00	106.10	4.1e+03	133	AW151370 xer75a07.x1 NCI_CGAP
gb_est25:AI267774	6.00	107.78	3.3e+03	105	AI267774 ap84h11.r1 TgrH_Tad	gb_est11:R25525	6.00	106.04	4.1e+03	134	R25525 EST00558 Unequalized
gb_est42:AO010782	6.00	107.72	3.3e+03	106	AO010782 AO010782 Schizosacchar	gb_est13:AA0739363	6.00	106.04	4.1e+03	134	AA0739363 j1473.seq.F Human f
gb_est7:W64053	6.00	107.52	3.4e+03	109	W64053 md77d11.r1 Soares mouse	gb_est16:AA565750	6.00	105.94	4.2e+03	136	AA565750 nj23h12.s1 NCI_CGAP
gb_est28:AI471311	6.00	107.52	3.4e+03	109	AI471311 ta14h03.x1 NCI_CGAP_Ly	gb_est17:AA611292	6.00	105.94	4.2e+03	136	AA611292 v171h07.r1 Knowles
gb_est45:AW259522	6.00	107.52	3.4e+03	109	AW259522 up28g09.x1 NCI_CGAP_Ma	gb_gss3:R28459	6.00	105.94	4.2e+03	136	R28459 F44L2TR IGF Arabidop
gb_est20:AA873029	6.00	107.39	3.5e+03	111	AA873029 ob11e10.s1 NCI_CGAP_Ki	gb_est14:AA420637	6.00	105.89	4.2e+03	137	AA420637 nc61f10.r1 NCI_CGAP

gb_est18:AA674007 - 6.00 105.89 4.2e+03 137 ! AA674007 vp70a10.r1 Knowles Sol
 gb_est18:AA687136 - 6.00 105.89 4.2e+03 137 ! AA687136 nv63e02.s1 NCI_CGAP_GD
 gb_est38:AA082321 + 6.00 105.89 4.2e+03 137 ! AA082321 xb65d10.x1 Soares_NFL
 gb_gss9:AA076792 + 6.00 105.89 4.2e+03 137 ! AO076792 CIT-HSP-236115.TR CIT-
 gb_est15:AA480572 + 6.00 105.83 4.2e+03 138 ! AA480572 nc70e07.r1 NCI_CGAP_PX
 gb_est19:AA775657 + 6.00 105.78 4.2e+03 139 ! AA775657 zf26d10.s1 Soares_feta
 gb_est15:AA243485 + 6.00 105.78 4.2e+03 139 ! AA243485 xm95h06.x1 NCI_CGAP_K1
 gb_est14:AA09493 + 6.00 105.73 4.2e+03 140 ! AA09493 EST01255 Mouse 7.5 dpd
 gb_est16:AA597233 + 6.00 105.73 4.2e+03 140 ! AA597233 v028c04.r1 Barstead mc
 gb_gss1:FR0022013 + 6.00 105.73 4.2e+03 140 ! AL014884 F.rubripes GSS sequenc
 gb_est21:AA948266 + 6.00 105.68 4.2e+03 141 ! AA948266 Oq34a05.s1 NCI_CGAP_GD
 gb_est15:AA483254 + 6.00 105.53 4.2e+03 142 ! AA483254 nf03f10.s1 NCI_CGAP_L4

seq_name: gb_est36:AI373988

seq_documentation_block: 283 bp mRNA EST 15-JAN-1999
 LOCUS AI373988
 DEFINITION SNOVAFCAP32B02SK Onchocerca volvulus adult female cDNA
 (SAW98MLW-OvAF) Onchocerca volvulus cDNA clone SNOVAFCAP32B02 5',
 mRNA sequence.

ACCESSION AI373988
 VERSION AI373988.1 GI:4160024

KEYWORDS EST.

SOURCE Onchocerca volvulus.

ORGANISM Onchocerca volvulus.

REFERENCE 1 (bases 1 to 283)
 Filarioidea; Onchocercidae; Onchocerca.

AUTHORS Lizotte-Waniewski, M. and Williams, S.A.

TITLE Genes expressed in adult female stage of Onchocerca volvulus

JOURNAL Unpublished (1998)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1797276.

Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pBluescript SK.

Location/Qualifiers

1..283

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone="SNOVAFCAP32B02"

/lab_host="X11-Blue MRF"

/sex="female"

/dev_stage="adult"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

xho I; Filarial nematode parasite of humans. Two adult

female worms of Onchocerca volvulus were isolated from

consenting patients and quick frozen. Adult female mRNA

was converted to double-stranded cDNA using reverse

transcriptase and oligo(dT) followed by RNase H and DNA

pol I. The library has 7 x 10E5 independent recombinants

and the average insert size is ~1100bp. The library was

constructed by Michelle Lizotte-Waniewski with worms

provided by Dr. Sara Lustigman. The library is available

from Dr. Steven A. Williams, email: genome@smith.edu."

63 a 86 c 74 g 60 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 52.00 Length: 52

Ratio: 1.000 Caps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AI373988

Align seg 1/1 to: AI373988 from: 1 to: 283

151 TTPLeuTTPProArgAsnProProileArgGluArgLysserileProTh 167
 |||||
 87 TGGTATGGCGGGGATCCACCAATTCGTGAACGCAAAAGTATCCAC 136
 |||||
 167 rSerTTPLeuLyileThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
 |||||
 137 CAGCTGGCTCAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGGCC 186
 |||||
 184 rGvettThraLaHlsValGlyPheProThrLeuArgLeuileArgTyrAla 200
 |||||
 187 GCATGACCGCCCATGTTGGCTTCCACACGCTGCGACTGATTCGCTATGCG 236
 |||||

201 MetGly 202

|||||

237 ATGGGT 242

seq_name: em_est20:AW017830

seq_documentation_block:

ID AW017830 standard; RNA; EST; 295 BP.

XX

AC AW017830;

XX

SV AW017830.1

XX

14-SEP-1999 (Rel. 61, Created)

DT 14-SEP-1999 (Rel. 61, Last updated, Ver)

XX

614064A09.x1 614 - root cDNA library f

DE mRNA sequence.

XX

EST.

XX

OS Zea mays

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

XX

[1]

RP 1-295

RA Walbot V.;

RT "Maize ESTs from various cDNA libraries sequenced at Stanford University";

RL Unpublished.

XX

MaizeDB; Probe/236176; p-std614064A09.

DR

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

SQ Sequence 295 BP; 80 A; 76 C; 76 G; 63 T; 0 other;

alignment_scores:
Quality: 25.00 Length: 25
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AW017830 ..

Align seg 1/1 to: AW017830 from: 1 to: 295

68 PheileProValGlnGlyValTyAlaAlaGlyArgLeuAspArgaspSe 84
|||||
146 TTCATCCCGGTTACGGGTGTTATGACAGCGAGTGCCTTGACCGCATAG 195

84 rGluGlyLeuValLeuThrAsn 92
|||||
196 CGAAGGTTGCTGCTGTGACCAAT 220

seq_name: gb_gss11:AQ212974

seq documentation_block:
LOCUS AQ212974 546 bp DNA GSS 18-SEP-1998
DEFINITION HS_3214_B2_E03_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3214 Col-6 Row-J, genomic survey sequence.
ACCESSION AQ212974
VERSION AQ212974.1 GI:3624175
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 546)
AUTHORS Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequencing-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT On Sep 10, 1998 this sequence version replaced gi:3554741.
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3214 row: J column: 6
Class: BAC end
High quality sequence stop: 546.

FEATURES

Location/Qualifiers
1..546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate-3214 Col-6 Row-J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 144 a 107 c 120 g 168 t 7 others
ORIGIN

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AQ212974/rev ..

Align seg 1/1 to reverse of: AQ212974 from: 1 to: 546

94 GlyValLeuGlnAlaArgLeuThrGlnPro 103
|||||
408 GGTGTGCTGCAGCGAGGCTGACACGCCN 379

seq_name: gb_est11:AA237292

seq documentation_block:
LOCUS AA237292 160 bp mRNA EST 03-MAR-1997
DEFINITION mw96d12.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:678551 5' similar to TR:G478984 G478984 PEROXISOMAL ENOYL HYDRATASE-LIKE PROTEIN. ; mRNA sequence.
ACCESSION AA237292
VERSION AA237292.1 GI:1861313
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 160)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilton,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404819.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:418255

Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 1.

FEATURES

Location/Qualifiers
1..160
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:678551"
/clone_lib="Soares mouse NML"
/tissue.type="Liver"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGCGCGGAATCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 36 a 40 c 41 g 43 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

```

US-09-252-691-7056 x AA237292/rev ..
Align seg 1/1 to reverse of: AA237292 from: 1 to: 160

83 aspSerGluclyLeuLeuValLeu 90
|||||
31 GACAGTGAAGGTCCTCTGTGACTT 8

seq_name: gb_est19:AA755409

seq_documentation_block: 232 bp mRNA EST 21-JAN-1998
LOCUS AA755409
DEFINITION VF53606.r1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
IMAGE:1124386 3' similar to SW:ECHE_RAT Q62651 PROBABLE PEROXISOMAL
ENOYL-COA HYDRATASE ; mRNA sequence.
ACCESSION AA755409
VERSION AA755409.1 GI:2802607
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 232)
AUTHORS Marra,M., Hillier,D., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900891.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:613722
High quality sequence stop: 224.
FEATURES
source
1. .232
Location/Qualifiers
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:1124386"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/Note="Organ: embryo; Vector: pBluescribe (modified);
Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
mRNA prepared from 13,500 2-cell stage embryos. Primer:
SalI(dT): 5'-CGGCGACCGCGACCGGTTTTTTTTTT-3'. cDNAs
were cloned into the MluI/SalI sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."
BASE COUNT 58 a 61 c 63 g 50 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AA755409/rev ..
Align seg 1/1 to reverse of: AA755409 from: 1 to: 232

83 aspSerGluclyLeuLeuValLeu 90
|||||
207 GACAGTGAAGGTCCTCTGTGACTT 184

seq_name: gb_est39:AV206648

seq_documentation_block: 248 bp mRNA EST 30-OCT-1999
LOCUS AV206648
DEFINITION AV206648 RIKEN full-length enriched, adult male testis Mus musculus
cDNA clone 170008L03 3', similar to X51703 Mouse mRNA for
ubiquitin, mRNA sequence.
ACCESSION AV206648
VERSION AV206648.1 GI:6147501
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 248)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N.,
Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3137194.
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@tc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitzunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
1. .248
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="170008L03"
/clone_lib="RIKEN full-length enriched, adult male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/Note="Site_1: XhoI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken

```

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTTTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGAGGCGCGCAATTAATTCGAGTTAATAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites."

BASE COUNT 75 a 60 c 41 g 72 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AV206648 ..

Align seg 1/1 to: AV206648 from: 1 to: 248

29 ArgGlnAlaThrArgAlaThrPro 36

9 CGCAGACACACAGAGACCCCC 32

seq_name: gb_est12:AA286664

seq_documentation_block: 249 bp mRNA EST 09-APR-1997
LOCUS AA286664
DEFINITION vb79h04.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone
IMAGE:763255 5' similar to TR:G478984 G478984 PEROXISOMAL ENOYL
HYDRATASE-LIKE PROTEIN. ;, mRNA sequence.

ACCESSION AA286664

VERSION AA286664.1 GI:1931759

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 249)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Apr 14, 1993 this sequence version replaced gi:504233.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:464175

Possible reversed clone: similarity on wrong strand

Seq primer: -28m3 rev2 Et from Amersham

High quality sequence stop: 157.

Location/Qualifiers

1. .249

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:763255"

/clone.lib="Soares mouse 3NME12 5"

/sex="unknown"

/tissue_type="fetus"

/dev_stage="12.5dpc total fetus"

FEATURES

Source

1. .256

/organism="Mus musculus"

/strain="B6D2 F1/J"

/db_xref="taxon:10090"

/clone="IMAGE:1111097"

/clone.lib="Knowles Solter mouse blastocyst B3"

/tissue_type="blastocyst"

/dev_stage="embryo (pre-implantation)"

/lab_host="DH10B"

/lab_host="DH10B"

/note="Organ: whole fetus; Vector: p7T3D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTCAAGTGGGAGCGCGCTATTTTTTTTTTTTTTTT
3'], on total mouse RNA [provided by Minoru Ko, Wayne
State Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 57 a 64 c 70 g 58 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AA286664/rev ..

Align seg 1/1 to reverse of: AA286664 from: 1 to: 249

83 AspSerGluGlyLeuValLeu 90

42 GACAGTGAAGGTCTTCGTACTT 19

seq_name: gb_est18:AA688759

seq_documentation_block:

LOCUS AA688759 256 bp mRNA EST 12-DEC-1997
DEFINITION vr07b09.r1 Knowles Solter mouse blastocyst B3 Mus musculus cDNA
clone IMAGE:1111097 5' similar to TR:Q45557 Q45557 ;, mRNA
sequence.

ACCESSION AA688759

VERSION AA688759.1 GI:2678188

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 256)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Nov 6, 1997 this sequence version replaced gi:932290.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:609285

High quality sequence stop: 165.

Location/Qualifiers

1. .256

/organism="Mus musculus"

/strain="B6D2 F1/J"

/db_xref="taxon:10090"

/clone="IMAGE:1111097"

/clone.lib="Knowles Solter mouse blastocyst B3"

/tissue_type="blastocyst"

/dev_stage="embryo (pre-implantation)"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pSPORT; Site_1: NotI; Site_2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(dT): 5'-CGTCACCGTCGACGCTTTT-3', cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies). Two different size selections: B1 (larger inserts) and B3."

BASE COUNT 49 a 69 c 77 g 61 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AA688759 ..

Align seg 1/1 to: AA688759 from: 1 to: 256

180 ArgGlnValArgMetThrAla 187
|||||
208 AGACAGGTGCGGAGGATGACAGCT 231

seq_name: gb_est1:X73795

seq_documentation_block:
LOCUS X73795 271 bp mRNA EST 13-NOV-1996
DEFINITION CHESTM092 Goat mammary gland Capra hircus cDNA, mRNA sequence.

ACCESSION X73795
VERSION X73795.1 GI:313488
KEYWORDS EST.

SOURCE goat.

ORGANISM

Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Caprinae; Capra.

REFERENCE 1 (bases 1 to 271)
AUTHORS Le Provost,F., Lepingle,A. and Martin,P.

TITLE A survey of the goat genome transcribed in the lactating mammary gland

JOURNAL Mamm. Genome 7 (9), 657-666 (1996)

MEDLINE 96359152

CONTACT: Le Provost,F.

LABORATOIRE de Genetique Biochimique

INRA-CRJ

78352 Jouy-en-Josas Cedex, FRANCE.

FEATURES

Source

1..271 Location/Qualifiers

/organism="Capra hircus"

/db_xref="taxon:9925"

/clone_lib="Goat mammary gland"

/note="Mammary gland of Capra hircus (goat) "

BASE COUNT 97 a 63 c 38 g 73 t

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X73795/rev ..

Align seg 1/1 to reverse of: X73795 from: 1 to: 271

42 ArgValIleLeuPheAsnLysPro 49
|||||
224 AGAGTAATCTCTTTTACAGCCA 201

seq_name: gb_est27:AI424418

seq_documentation_block:

LOCUS AI424418 291 bp mRNA EST 28-MAR-1999
DEFINITION te90f02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2093979 3', similar to gb:S37431 LAMININ RECEPTOR (HUMAN);, mRNA sequence.

ACCESSION AI424418

VERSION AI424418.1 GI:4270349

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 291)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Apr 21, 1998 this sequence version replaced gi:3072089.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Insert Length: 457 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 1.

FEATURES

Location/Qualifiers

1..291

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2093979"

/clone_lib="NCI_CGAP_Pr28"

/sex="male"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the

normalized library NCI_CGAP_Pr22 was prepared, and ss

circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from a pool

of 5,000 clones made from the same library (cloneids

985608-986759, 1101192-1101959, and 1217928-1220615).

Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 68 a 76 c 80 g 67 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AI424418

Align seg 1/1 to: AI424418 from: 1 to: 291

59 GluAlaClyArgSerThrLeuLys 66

|||||

106 GAGCGTGGCGCTCCACTCTTAAG 129

seq_name: gb_est42:AV309318

seq_documentation_block:

```

LOCUS       AV309318             304 bp      mRNA           11-NOV-1999
DEFINITION   AV309318 RIKEN full-length enriched, 8 days embryo Mus musculus
              cDNA clone 5730578K03 3', mRNA sequence.
ACCESSION    AV309318
VERSION      AV309318.1 GI:6362353
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 304)
AUTHORS      Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
              Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
              Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
              Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
              Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
              Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
              Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
              Suzuki,H., Suzuki,H., Takahashi,F., Tateo,M., Tominaga,N.,
              Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
              Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
              RIKEN Mouse ESTs (Konno,H., et al.)
              Unpublished (1999)
TITLE        On May 18, 1998 this sequence version replaced gi:3138303.
JOURNAL      Contact: Yoshihide Hayashizaki
              Genome Exploration Research Group, Life Science Tsukuba Center,
              The Institute of Physical and Chemical Research (RIKEN), Genomic
              Sciences Center
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: +81-298-36-9013
              Fax: +81-298-36-9098
              Email: genome.res@rtc.riken.go.jp,
              URL:http://genome.rtc.riken.go.jp/
              Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
              Matsunura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
              Hayashizaki,Y.
              Transcriptional sequencing: A method for DNA sequencing using RNA
              polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
              Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
              Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
              Okazaki,Y. and Hayashizaki,Y.
              Automated filtration-based high-throughput plasmid preparation
              system. Genome Res. 9 (5), 463-470 (1999)
              Carninci,P. and Hayashizaki,Y.
              High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
              19-44 (1999)
              Please visit our web site (http://genome.rtc.riken.go.jp) for
              further details.
FEATURES     Location/Qualifiers
              1..304
                 /organism="Mus musculus"
                 /strain="C57BL/6J"
                 /db_xref="taxon:10090"
                 /clone="5730578K03"
                 /clone_lib="RIKEN full-length enriched, 8 days embryo"
                 /sex="mixed"
                 /dev_stage="8 days embryo"
                 /lab_host="DH10B"
                 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                 prepared and sequenced in Mouse Genome Encyclopedia
                 Project of Genome Exploration Research Group in Riken
                 Genomic Sciences Center and Genome Science Laboratory in
                 RIKEN, Division of Experimental Animal Research in Riken
                 contributed to prepare mouse tissues. 1st strand cDNA was
                 primed with a primer [5'
                 GAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was
                 prepared by using trehalose thermo-activated reverse
                 transcriptase and subsequently enriched for full-length by
                 cap-trapper. cDNA went through one round of subtraction to
                 Rot = 100.0 Second strand cDNA was prepared with the
                 primer adapter of sequence [5'
                 GAGAGAGAGATCTCGAGTAAATTAATTAATCCCCCCCCCC 3']. cDNA

```

```

was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FIC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT   93 a      82 c      61 g      68 t
ORIGIN
alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:        0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-252-691-7056 x AV309318/rev ..
Align seg 1/1 to reverse of: AV309318 from: 1 to: 304
83 AspSerGluGlyLeuLeuValLeu 90
|||||
45 GACTCTGAGGCGCTTACTGCTGTTG 22
seq_name: gb_est22:AI047546
seq documentation_block:
LOCUS       AI047546             338 bp      mRNA           08-JUL-1998
DEFINITION   un08005.r1 Soares mouse urogenital ridge NMUR Mus musculus cDNA
              clone IMAGE:1763985 5' similar to TR:035459 O35459 ECH1P.;, mRNa
              sequence.
ACCESSION    AI047546
VERSION      AI047546.1 GI:3295833
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 338)
AUTHORS      Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Stepto,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
              The WashU-HMI Mouse EST Project
              Unpublished (1996)
              On Jan 14, 1998 this sequence version replaced gi:2754372.
              Contact: Marra M/Mouse EST Project
              WashU-HMI Mouse EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LLNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:963509
              Possible reversed clone: similarity on wrong strand
              Seq primer: -28mi3 rev2 ET from Amersham
              High quality sequence stop: 107.
FEATURES     Location/Qualifiers
              1..338
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /clone="IMAGE:1763985"
                 /clone_lib="Soares mouse urogenital ridge NMUR"
                 /sex="equal ratio of male:female"
                 /tissue_type="urogenital ridge (embryonic)"
                 /dev_stage="fetal, mixture of 11.5 and 12.5 dpc"
                 /lab_host="DH10B"
                 /note="Organ: gonad; Vector: pT7T3D-Pac (Pharmacia) with a
                 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                 strand cDNA was primed with a Not I - oligo(dT) primer [5'
                 TGTTACCATCTGAAGTGGAGCGCGCCGCTCTTTTCTTTTCTTTTCTTTT
                 T 3']; double-stranded cDNA was ligated to Eco RI
                 adaptors (Pharmacia), digested with Not I and cloned into

```

the Not I and Eco RI sites of the modified pT73 vector.
Library went through two rounds of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 74 a 86 c 100 g 78 t

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AI047546/rev ..
Align seg 1/1 to reverse of: AI047546 from: 1 to: 338

83 AspSerGluGlyLeuValLeu 90
26 GACAGTGAAGGCTCTCTGGTACTT 3

seq_name: gb_est26:AI329557

seq_documentation_block:
LOCUS AI329557 350 bp mRNA EST 28-DEC-1998
DEFINITION b6c08ne.fl Neurospora crassa evening cDNA library Neurospora crassa
CDNA clone b6c08ne 5', mRNA sequence.

ACCESSION AI329557
VERSION AI329557.1 GI:4066116
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa

Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae;

REFERENCE
AUTHORS Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
TITLE Two Neurospora crassa EST Databases
JOURNAL Unpublished (1998)
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 135.

Location/Qualifiers

FEATURES

source

1..350
/organism="Neurospora crassa"
/strain="Strain 30-7 (bd; A)"
/db_xref="taxon:5141"
/clone="b6c08ne"
/clone_lib="Neurospora crassa evening cDNA library"
/tissue_type="tissue harvested following 22hr growth in
dark"
/note="Vector: pBlueScript SK-; Site_1: XbaI; Site_2:
EcoRI; See: Bell-Perdersen, D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3'
end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT 96 a 92 c 78 g 84 t

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AI329557

Align seg 1/1 to: AI329557 from: 1 to: 350

191 PheProThrLeuArgLeuLeuArg 198
|||||
209 TTCCCAACTCTACGGCTTATCCGA 232

seq_name: gb_est38:AW047543

seq_documentation_block:
LOCUS AW047543 357 bp mRNA EST 18-SEP-1999
DEFINITION UI-M-BH1-ama-f-02-0-UI.s1 NIH_BMAP_M_S2 Mus musculus CDNA clone
UI-M-BH1-ama-f-02-0-UI 3', mRNA sequence.

ACCESSION AW047543
VERSION AW047543.1 GI:5908072
KEYWORDS EST.
SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL
MEDLINE 9704477
COMMENT On Jun 22, 1998 this sequence version replaced gi.3246699.

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643 USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized brain stems library cDNA library preparation: M.B.
Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA=yes

Location/Qualifiers

FEATURES

source

1..357
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH1-ama-f-02-0-UI"
/clone_lib="NIH_BMAP_M_S2"
/dev_stage="27-32 days"
/lab_host="PH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S2 library is a subtracted library derived from
NIH_BMAP_M_S1, which in turn is a subtracted library
derived from a mixture of normalized libraries from ten
regions of the mouse brain (cerebellum, brain stems,
olfactory bulbs, hypothalamus, cortex, amygdala, basal
ganglia, pineal gland, striatum, hippocampus). The driver
used for subtraction consisted of a pool of 5,000 clones
from the NIH_BMAP_M_S1 library and a pool of 2,000 clones
obtained from non-normalized and normalized mouse brain
spinal cord libraries.
TAG LIB=NIH_BMAP_M_S2
TAG TISSUE=brain-stems
TAG_SEQ=TCATG"

BASE COUNT 83 a 86 g 93 t 1 others

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x AW047543 ..

Align seg 1/1 to: AW047543 from: 1 to: 357

83 AspSerGluGlyLeuLeuValLeu 90

|||||

154 GACAGTGAAGGCTCTTCGTACTT 177

seq_name: gb_est1:T52062

seq_documentation_block:

LOCUS T52062 374 bp mRNA EST 06-FEB-1995
DEFINITION YB29d04.r1 Stratagene fetal spleen (#937205) Homo sapiens cDNA
clone IMAGE:72583 5' similar to contains MER6 repetitive element,
mRNA sequence.

ACCESSION T52062
VERSION T52062.1 GI:653922

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 374)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)

MEDLINE

9704478

COMMENT

Contact: Wilson RK

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 456

High quality sequence stops: 226 Source: IMAGE Consortium, LLNL This

clone is available royalty-free through LLNL; contact the IMAGE

Consortium (info@image.llnl.gov) for further information.

Insert Length: 456 Std Error: 0.00

Seq primer: M13Rpl

High quality sequence stop: 226.

FEATURES

Source

1..374
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:494248"
/db_xref="taxon:9606"
/clone="IMAGE:72583"
/clone_lib="Stratagene fetal spleen (#937205)"
/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: spleen; Vector: pBluescript SK-; Site: 1:
ECORI; Site: 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled spleens. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAG
3' -3' adaptor sequence: 5' CTCAGTTTTTTTTTTT 3'

BASE COUNT

ORIGIN

89 a 88 c 109 g 85 t

3 others

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x T52062 ..

Align seg 1/1 to: T52062 from: 1 to: 374

59 GluAlaGlyArgSerThrLeuLys 66

|||||

218 GAGCGGGCAGATCCACCTTGAAG 241

ID W94884 standard; protein; 88 AA.
AC W94884;
DT 25-MAY-1999 (first entry)
DE Colicin V cleaved peptide sequence.
KW Enterocin 900; brochocin-C; bacterial growth; microorganism; inhibit;
KW food; meat; fermentation; bacteriocin; mastitis; secretory vector; lcaC;
KW leucocin A; lcaD; enzyme; probiotic; L. gelidium; Listeria; Lactobacillus;
KW sulphide-producing; spoilage; sensory property; smell; taste; infection;
KW ground beef; silage; Salmonella; poultry; intestine; therapeutic; colour;
KW mucosal tissue; vacuum-packed; storage; colicin V.
OS Escherichia coli.
FH Key Location/Qualifiers
FT Disulfide_bond 76..87 /note= "disulphide bridge"
FT W09902555-A1.
PN 21-JAN-1999.
PD 09-JUL-1998; U14331.
PF 05-SEP-1997; US-924629.
PR 09-JUL-1997; US-052072.
PA (UYAL-) UNIV ALBERTA.
PI Franz C, Greer GG, Leisner JJ, McCormick JK, McMullen LM,
PI Poon A, Stiles ME, Van Belkum MJ, Vederas JC, Worobo RJ,
PI Worobo RW;
PI WPI; 99-120779/10.
DR New bacteriocins enterocin 900 and brochocin-C - useful as food
PT preservatives and therapeutic antimicrobials
PT Disclosure; Page 138-139; 182pp; English.
CC The invention relates to peptides enterocin 900, brochocin-C that inhibit
CC bacterial growth. Microorganisms expressing the above peptides are used
CC to inhibit bacterial growth in foods (specifically meat), live animals
CC (applied topically), food preparation areas and fermentation vessels. The
CC bacteriocins enterocin 900, brochocin-C may be used directly, specifically
CC to inhibit bacteria that cause mastitis. Antibodies specific for the
CC peptides are used to detect expression of bacteriocins in cells. Secretory
CC vectors (containing sequences for a promoter, a leucocin A processing
CC peptide, a heterologous peptide, lcaC and lcaD) are used to secrete
CC bacteriocins, enzymes or other proteins, e.g. for use in food production
CC and as probiotics. L. gelidium, or other leucocin A-expressing bacteria,
CC inhibit the growth of Listeria and the sulphide-producing spoilage
CC organism Lactobacillus sake; they also improve the sensory properties
CC (smell, taste and colour) of meat, specifically ground beef. L. gelidium
CC can also be used in preparation of animal feeds (silage), as probiotic,
CC to control Salmonella in poultry intestines and therapeutically against
CC mucosal tissue infections. Treatment with bacteriocins provides vacuum-
CC packed meat products with predictable and longer storage life. The
CC secretion vectors can express several bacteriocins for broader spectrum
CC of activity, e.g. tailored for a particular target organism.
SQ Sequence 88 AA;

Query Match 2.7%; Score 6; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. NO. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 75 YAAAGRL 80
Db 70 YAAAGRL 75

Search completed: June 10, 2000, 11:27:58
Job time: 2488 sec

THIS PAGE BLANK (USPTO)